AGARD-CP-596

AGARD-CP-596

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT
7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARD CONFERENCE PROCEEDINGS 596

Audio Effectiveness in Aviation

(l'Efficacité des communications vocales en aéronautique)

Papers presented at the Aerospace Medical Panel Symposium on "Audio Effectiveness in Aviation" held in Copenhagen, Denmark, 7-10 October 1996.

NORTH ATLANTIC TREATY ORGANIZATION

Published June 1997

Distribution and Availability on Back Cover

AGARD-CP-596

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE, 92200 NEUILLY-SUR-SEINE, FRANCE

AGARD CONFERENCE PROCEEDINDGS

Audio Effectiveness in Aviation

(l'Efficacité des communications vocales en aéronautique)

Papers presented at the Aerospace Medical Panel Symposium on "Audio Effectiveness in Aviation" held in Copenhagen, Denmark, 7-10 October 1996.

North Atlantic Treaty Organization

Organisation du Traité de l'Atlantique Nord

The Mission of AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

— Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;

— Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);

— Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

— Improving the co-operation among member nations in aerospace research and development;

— Exchange of scientific and technical information;

— Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

— Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published June 1997

ISBN 92-836-0043-6

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Audio Effectiveness in Aviation

(AGARD CP-596)

Executive Summary

The Aerospace Medical Panel (AMP) of the Advisory Group for Aerospace Research and Development (AGARD) held a Symposium entitled "Audio Effectiveness in Aviation", in Copenhagen, Denmark, 7-11 October 1996. The Symposium was held in order to address concerns that, while effective voice communications and aural signals are important in military and civil aviation, their implementations are often less than satisfactory in modern aircraft. Factors that influence this are:

(a) in many cases audio communications systems in aircraft are based on design concepts that are dated and do not take advantage of recent advances in the area;

(b) the noise environments in which the aviator performs often cause acoustic interference with attempts to communicate by means of auditory channels;

(c) prolonged exposure to those same noise environments causes temporary and even permanent hearing impairment to the operator;

(d) audio signals used as warnings, cautions and advisories are non-standardized and not optimally designed;

(e) often audio displays are designed without adequate consideration being given to how they will be integrated into the aircraft systems within which they will function.

The Symposium identified the above-mentioned factors and covered a number of topics that will provide military benefits.

These benefits include:

- Signal processing technologies that will mitigate the effects of the operational noise environment and enhance communications in that environment
- New sound attenuation materials that will improve passive hearing protection devices
- New voice communication tests that will more effectively predict the performance of communication systems in operational environments
- New audio display technologies that will provide spatial audio information to the operator thereby enhancing communications, providing threat warnings, and improving situational awareness
- New tests that allow the operator's emotional state to be inferred from his/her speech
- Guidelines for enhancing the effectiveness and utility of voice input interfaces in cockpit applications.

L'efficacité des communications vocales en aéronautique

(AGARD CP-596)

Synthèse

Le Panel de médecine aérospatiale de l'AGARD (AMP) a organisé un symposium sur "L'efficacité des communications vocales en aéronautique" à Copenhague, au Danemark, du 7 au 11 octobre 1996. Le symposium s'est donné pour objectif de déterminer pourquoi la mise en œuvre des communications vocales et des signaux sonores dans les aéronefs modernes, qui est considéré comme un sujet important par les autorités civiles et militaires, laisse si souvent à désirer. Les facteurs contribuant à cet état de fait sont les suivants:

(a) Dans beaucoup de cas les systèmes de communication vocale de bord sont basés sur des concepts démodés qui ne bénéficient pas des dernières avancées réalisées dans le domaine;

(b) les environnements sonores dans lesquels l'aviateur opère posent souvent des problèmes d'interférence acoustique dans les communications par voie auditive;

(c) l'exposition prolongée à ces mêmes environnements sonores peut occasionner la détérioration temporaire voire permanente de l'ouie de l'opérateur;

(d) les signaux sonores utilisés en tant qu'avertissements, mises en garde et avis sont non-standardisés et leur conception n'est pas optimisée;

(e) la conception des systèmes de présentation d'informations vocales tient rarement compte de l'intégration des équipements dans l'avionique existante.

Le symposium a examiné les facteurs susmentionnés et a traité un certain nombre de questions susceptibles de fournir des avantages aux militaires.

Ces avantages comprennent:

- des technologies de traitement du signal qui permettront d'atténuer les effets de l'environnement sonore opérationnel et d'améliorer la qualité des communications en cet environnement
- de nouveaux matériaux absorbant le son qui permettront d'améliorer les performances des dispositifs de protection de l'ouie passive
- de nouveaux tests pour les communications vocales qui permettront de prévoir avec plus de précision les performances des systèmes de communications en environnement opérationnel
- de nouvelles technologies de présentation d'informations vocales qui permettront de transmettre à l'opérateur des informations spatiales sous forme vocale, et de ce fait, d'améliorer les communications, de fournir des avertissementes de la menace et de disposer d'une meilleure appréciation de la situation
- de nouveaux tests permettant de déterminer l'état émotif de l'opérateur à partir de l'inflexion de sa voix
- des directives pour l'amélioration de l'efficacité et de l'utilité des interfaces des commandes vocales du poste de pilotage.

Contents

NOISE CONTROL

SESSION II: PASSIVE TECHNIQUES

SESSION III: ACTIVE TECHNIQUES

SESSION IV: APPLICATIONS

SPEECH TECHNOLOGY

SESSION V: COMMUNICATION IN STRESSFUL ENVIRONMENTS

SESSION VI: VOICE CONTROL

† Paper not available at time of printing.

Preface

The auditory channel is second in importance only to the visual channel in the presentation of information in the flight environment. Despite that fact, serious shortcomings exist in the representation and effectiveness of auditory information in the cockpit. The shortcomings that exist often reflect a failure either to take advantage of auditory research that exists or to prepare to incorporate audio technologies that are in various states of development. The purpose of this Symposium was to address some of the problems that impede the application of audio technologies in the operational environment and to acquaint the AGARD community with current research and technologies under development that have the potential of increasing the effectiveness of the operator.

The papers presented addressed research and the development and applications of technologies in the areas of:

(a) Audio Displays;

(b) Passive and Active Noise Control;

(c) Communications in Stressful Environment;

(d) Voice Control.

These proceedings will be of interest to those concerned with the health and safety of personnel in air and support operations; those concerned with the presentation of information by means of the auditory channel and the input of information to machines by way of speech; and the aerospace scientist wanting a review of relevant research in the fields of hearing protection, audio displays, voice communications, and voice control.

Topics addressed during this Symposium were:

- Audio Displays
- Noise Control - Passive Technique
- Noise Control - Active Technique
- Noise Control - Applications
- Communication in Stressful Environment
- Voice Control

Aerospace Medical Panel

Chairman: Dr P. VANDENBOSCH
Loriesstraat, 44
B-1500 Halle, Belgium

Deputy Chairman: LtCol A. ALNAES
Oslo Military Clinic
Oslo Mil/Akershus
N-0015 Oslo, Norway

TECHNICAL PROGRAMME COMMITTEE

Co-Chairmen

Dr R. BURTON
Armstrong Laboratory AL/CN
2509 Kennedy Drive
Brooks AFB, TX 78235-5118
USA

Colonel D.F. SHANAHAN
U.S. Army Aeromedical Research Laboratory
Fort Rucker, AL 36362-0577
USA

Members

Capt. I. DIAMANTOPOULOS
HAF 251 Gen. Hospital
P. Kanellopoulou & Messogion Ave.
11525 Athens
Greece

Dr A.R. LEGER
Sextant Avionique
Rue Toussaint Catros
BP 91
33166 St Médard en Jalles Cedex
France

Dr E. SCHROEDER
Lemschesvej 20
DK - 2900 Hellerup
Denmark

Dr G.M. ROOD
Manager Mission Management Dept.
Defence Research Agency
Room 204, Bldg R177
Farnborough, Hants GU14 6TD
United Kingdom

LOCAL COORDINATOR

Dr E. SCHROEDER
Lemschesvej 20
DK - 2900 Hellerup
Denmark

PANEL EXECUTIVE

Mail from Europe & Canada:
Major R. POISSON, CF
AGARD/NATO
7, rue Ancelle
92200 Neuilly-sur-Seine

Mail from USA:
AGARD/NATO/AMP
PSC 116
APO AE 09777

Telephone: (33) (0)1 55 61 22 60/62
Telex: 610176F
Telefax: (33) (0)1 55 61 22 99/98

TECHNICAL EVALUATION REPORT

2610 Seventh Street
ARMSTRONG LABORATORY, AL/CFBA
WRIGHT-PATTERSON AFB OHIO 45433-7901
USA

Thomas J. Moore, Ph.D.

1. INTRODUCTION

The AGARD Aerospace Medical Panel held a Symposium on "Audio Effectiveness in Aviation" in Copenhagen, Denmark, 7-11 October 1996. Thirty-four contributed papers were presented along with an invited presentation, a key-note address, three overview presentations of relevant technology areas and a summary presentation. Papers represented contributions by authors from eight NATO countries and Australia.

2. THEME

The theme of the Symposium was that effective voice communications and aural signals are vital for military, and civil aviation. Despite that fact, audio communications are often less than desired in modern aircraft.

This shortfall in capability is due to a number of factors. The audio communications systems, in many cases, are based on design concepts that are many years old and do not take advantage of research that has advanced our understanding of the area. The noise environments in which the aviator typically performs can often cause acoustic interference with attempts to communicate via auditory channels. In addition, these noise environments may cause physiological changes in the auditory system that result in temporary, and eventually permanent loss of hearing, further impeding communications. A lack of standardization of audio signals used as cautions, warnings, and advisories to air crew often results in confusion during time-critical operations and emergencies.

Finally, there are many obvious applications of voice input/voice output technologies in aviation. Despite the considerable amount of research over the last 20 years devoted to applying voice technologies to aviation, the successful integration of voice input/voice output into the cockpit has not yet been accomplished.

3. PURPOSE AND SCOPE

The stated purpose of the Symposium was to present current basic and applied research efforts to address factors that limit the effectiveness of audio communications in aviation and related operational environments. Papers were solicited and received describing basic and applied studies on topics such as:

- Measures of Communications
- Auditory Displays
- Voice Input/Voice Output
- Active and Passive Technologies for Hearing Protection
- Aviator Hearing Requirements for Communications
- Technology Integration

4. SYMPOSIUM PROGRAM

After the Opening Remarks by the Program Committee Chairman, Dr. R. R. Burton of the USA, the Symposium started with an historical review of Auditory Research in Denmark by Dr. S. Vesterhauge (DE). This was followed by the keynote address on the topic of the audio environment experienced in aircraft by Dr. G. M. Rood (UK). The Symposium was organized along the lines of three general technology areas, each area

consisting of 1 to 3 technical sessions. The sessions and their chairmen were as follows in chronological order:

Session I Audio Displays
Chairmen: Dr. A. R. Leger (FR)
Dr. R. R. Burton (USA)
Session II Noise Control: Passive Techniques
Chairmen: Dr. G. M. Rood (UK)
Dr. T. J. Moore (USA)
Session III Noise Control: Active Techniques
Chairmen: Dr. T. J. Moore, (USA)
Dr. G. M. Rood (UK)
Session IV Noise Control: Applications
Chairmen: Col D. F. Shanahan (USA)
Dr. G. M. Rood (UK)
Session V Speech Technology: Communications In Stressful Environments
Chairmen: Dr. R. D. Patterson (UK)
Capt I. Diamantopoulos (GR)
Session VI Speech Technology: Voice Control
Chairmen: Dr. A. R. Leger (FR)
Col D. F. Shanahan (USA)

Each of these general technology areas was initiated by an overview presentation by a technical expert who provided a comprehensive view of the state-of-the-art in the technical area and provided an overall context for the presentations that were to follow in the technical sessions under that area. These experts and the areas they provided overviews of were respectively:

a. Mr. R. L. McKinley (USA) - Audio Displays
b. Dr. A. Dancer (FR) - Noise Control
c. Dr. H. J. M. Steeneken (NE) and Mr. L. Gagnon (CA) - Speech Technology

5. TECHNICAL EVALUATION

In his key note address, Dr. Rood provided a survey of noise levels in both historical and current aircraft. He noted that in modern aircraft factors such as advances in propulsion technology and operational tactics (e.g., high-speed, low-level flight) have resulted in the noise problem in aircraft being even more severe than in the past. Dr. Rood then provided a survey of what the implications of this noise environment were for areas such as hearing damage risk and interference with aural communications (both speech and non-speech). He concluded by noting that existing and emerging audio technologies hold promise of mitigating the problems and allowing the enhancement of crew performance by removing existing impediments to the use of the auditory channel in operational environments.

5.1 Audio Displays

This session consisted of papers addressing the advantages of the use of spatial audio information for target detection and communication enhancement, as well as the criteria to be used in designing auditory warnings, cautions and advisories for aircraft application. After an overview of the area by Mr. McKinley that surveyed the present state-of-the-field in the presentation of spatial audio information via earphones, a series of papers ranging from basic laboratory studies, through simulator studies and actual flight demonstrations provided data that clearly demonstrated the potential of spatial auditory information to enhance target detection and communications effectiveness. Dr. B. Elias (paper #1) presented laboratory data that dynamic spatial auditory cues (presented over loud speakers) providing information regarding position and velocity of dynamic visual targets prior to their appearance on a visual display significantly reduced response times in a visual search task. This synergy between spatial auditory information and visual information in acquiring visual targets was also demonstrated in data acquired in a flight simulation environment using 3-D audio synthesizers which, when coupled with a head tracker, provide spatial audio information over headphones (Courneau, et al, paper #4, Bronkhorst and Veltman, paper #5). McKinley, et al, (paper #6) then provided flight demonstration data (using a

TAV-8B Harrier and 3-D audio) showing that in this demanding flight environment the advantages of spatial audio information in target acquisition were also found, and that in addition, the perception of multiple simultaneous voice communications was enhanced. Finally, paper #2 (Gilkey and Simpson) in this session provided laboratory data on the limits of the human's ability to localize sounds and speech signals in noise when the subject's head is fixed in position and discussed the implications of these performance limitations for the design of auditory displays.

A series of papers by James (paper #7), Patterson and Datta (paper #8), and Martin, et al (paper #9), dealt with the design of standardized audio warning signals by the UK's Defence Research Agency (DRA). James detailed the research conducted to develop a set of design guidelines for audio warning signals. Patterson and Datta discussed efforts to increase the frequency range of the existing set of sounds to enhance their localizability, while preserving their distinctiveness and recognizability, as well as work to develop a new set of threat warning sounds. Martin, et al, reported laboratory data on the ability to localize the existing set of warning sounds. Their data indicate that the temporal structure rather than the frequency content of the existing sound warning set may have the greater influence on localizability. All these papers note that due to the likelihood of the introduction of 3-D audio systems in future cockpits, the question of designing audio warnings to be highly localizable in noise is one that needs to be addressed.

In this session, paper #3, which was to discuss aspects of aircrew behavior in relation to audio-visual warning systems during periods of high work-load was not presented.

5.2 Noise Control

The overview presentation for this area was presented by Dr. A. Dancer. Dr. Dancer reviewed the state of passive and active techniques for noise control and provided the context within which the papers in the subsequent sessions on passive and active techniques and their applications were presented.

5.2.1 Passive Techniques

One paper in this session (#10) actually was concerned with the evaluation of active attenuation devices and will be dealt with in the next section. Also, one paper from the session on active techniques (#14) was more suited for this section. In the session on passive techniques, two papers (Reynaud, et al, paper #11, and Mozo and Ribera, paper #14) dealt with the use of passive hearing protectors (ear-plugs) fitted with miniature audio transducers for communications in operational environments. Reynaud, et al, examined the suitability of such devices when used in high performance aircraft. Potential problems were identified and practical solutions proposed. Mozo and Ribera evaluated the suitability of such devices for use in rotary-winged aircraft. Field test results reported comparing the performance of the ear-plug communication device with two Active Noise Reduction (ANR) systems favored the ear-plug device. This paper provoked spirited discussion among the attendees, with the concern being expressed that the actual noise attenuation that would be provided in operational use by the ear-plug device would be less than that provided by ANR because of the variability experienced in fitting of ear-plugs and their use in the operational environment. There was sufficient interest in and differing viewpoints on this question that time was provided in the final day of the Symposium to continue the discussion. Aside from a consensus that a need existed for a definitive controlled study, no agreement was reached between the proponents of the opposing viewpoints.

Another paper presented in this session dealt with the effect of the thickness of the frame of eyeglasses on the noise attenuation effectiveness of head sets and helmets (Rose, et al, paper #12). The thicker the frame, the more the integrity of the seal of the ear-cup was compromised and the less noise attenuation provided. Paper #13 (Thomas, et al) described a program of study to develop a material that could be used to develop a circumaural hearing protector that would be more effective in attenuation of low frequency noise than existing earcups. Data from a prototype design was presented.

5.2.2 Active Techniques

A number of papers at this Symposium dealt with the measurement of the attenuation effectiveness and speech intelligibility of commercially available ANR devices (Buck, et al, paper #10; Crabtree, paper #15; Pellieux, et al, paper #16; Wagstaff and Woxen, paper #17; Steeneken and Verhave, paper #18). In general, the results were consistent in finding that commercially available ANR headsets can differ considerably in such areas as attenuation effectiveness, tendency to overload, tendency to become unstable, and the existence of noise augmentation in the mid-frequencies when the ANR circuitry was on. A key problem that was noted often in discussion and mentioned explicitly by Steeneken and Verhave (paper #18), is the necessity of adapting a standardized method of evaluation of ANR systems so that results across laboratories can be compared meaningfully.

Two papers reported on attempts to solve stability and overload problems in existing analog ANR systems by developing hybrid analog/digital adaptive ANR systems (Pan and Brammer, paper #19, and Darlington and Rood, paper #23). Both papers provide evidence that ANR technology can be advanced to provide greater performance than contemporary commercially available systems.

5.2.3 Applications

In this session the application of ANR technology to improve the performance of the hearing impaired in high performance aircraft (McKinley, et al, paper #21), to enhance communications and hearing protection in armored vehicles (Anderson and Garinther, paper #20) and helicopters (Simpson and King, paper #22) was reported. McKinley, et al, reported on a demonstration of a specially modified ANR headset for use by an individual with a severe hearing loss. Anderson and Garinther discussed the experience of the U.S. Army in fielding a new armored vehicle intercom system that incorporates ANR technology. They report significant gains in noise attenuation, speech intelligibility and comfort compared to the previous system. Simpson and King reported on a study comparing the noise attenuation, speech intelligibility, perceived attention demand, and perceived operational suitability over the standard crew helmet for ANR and an ear plug communication device (paper #14) when used in helicopter operations. Their study indicated that ANR would appear to be the preferred solution based on better attenuation at low frequencies and high aircrew ratings.

5.3 Speech Technology

The overview for this area was given by Dr. H. J. M. Steeneken and Mr. L. Gagnon. They addressed the question of the relevance of speech and language technology in the military. Their paper reported on the coordinated speech technology research of nine NATO countries as represented by the activities of the NATO research study group on speech processing.

5.3.1 Communication in Stressful Environments

This session addressed the question of measurement of speech intelligibility in the aviation environment. Vesterhauge, et al

(paper #24), described a program that has been initiated in the Danish Air Force. Aircraft noise levels and spectra at various crew positions are measured and tape recorded. These recordings are then used in a ground-based test environment to assess the intelligibility of tape recorded air-traffic-control messages for individual crew members suffering from hearing impairment in order to assess their fitness for flight duty. Hanschke (paper #25) advocated the use of an accepted speech discrimination test rather than audiometric measures to determine the ability of aviators to safely operate in operational noise environments. Wagstaff (paper #26) investigated the combined effects of noise and altitude on hearing function. He reported a substantial increase in speech intelligibility in noise due to altitude. He explains these results by noting that with increased altitude, there would be a reduction in environmental noise due to decreased air density. The speech signal within the communication system would not be similarly reduced, resulting in an improved signal-to-noise ratio and thus increased intelligibility.

Nixon, et al (paper #27) reported on differences between male and female speech in different operational noise environments, as well as when the speech was processed by two types of vocoders and two different automatic speech recognition systems. The only statistically significant difference they found was that the intelligibility of female speech was poorer in high noise environments. The use of ANR technology and modern noise canceling microphones increased the intelligibility of the females' speech to where it was equivalent to that of the males. Finally, McKinley, et al, (paper #29) presented a Voice Communications Effectiveness Test that is a methodology and metric for relating voice communications performance to the effective completion of tasks with varying complexity, criticality, and time constraints. This test attempts to address the question of voice communications effectiveness, i.e., how much information is being communicated within the constraints of a specific operational scenario, rather than what percentage of a list of words is correctly identified.

In this session, paper #28, which was to also discuss the development of more reliable and valid measures of communications performance was not presented.

5.3.2 Voice Control

This session addressed the application of voice input devices in the cockpit environment and environmental influences that affect the performance of such systems. Allerhand and Patterson (paper #30) reported an application of a computational auditory model to measure vocal agitation in speech automatically. Such a method would provide a non-invasive method of determining whether an operator is under severe stress and has long been a goal of the speech research community. It is conceivable that if such an objective measure of emotional stress could be developed, it could also be used to adapt the performance of speech recognition devices to take into account variations in the operator's voice characteristics due to emotional stress. Rogers and Rood (paper #31) report on the intelligibility and user acceptability of speech processed over a communication link including a Linear Predictive Coder, in the presence of helicopter noise environments. Their study shows that with digital pre-processing of the speech signals the intelligibility of vocoded speed can be significantly improved; however, the user acceptability measures for vocoded speed remain very low relative to clear speech.

A series of papers in this session address the evaluation of the performance of automatic speech recognition systems in an F-16 simulator (Steeneken and Pijpers, paper #33), a Tornado strike aircraft (South, paper #34), an OV-10 aircraft (Williamson, paper #35) and a Gazelle helicopter and Alpha-Jet fighter (Cordonnier, et al, paper #36). The results from these studies have shown, as

many earlier studies have also, that although adverse environmental factors, such as noise, effects of oxygen mask, effects of pulling G, buffeting, etc., all affect the performance of the automatic speech recognition system, there is sufficient benefit to be derived from making such systems work in the flight environment that engineering and training solutions to these problems continue to be pursued.

The final paper in this session by Cook, et al (paper #37) presents data that demonstrate that the use of speech interfaces results in degraded performance on tasks requiring processing of visual information and problems in retrieving from memory auditory information. Unfortunately, the lack of adequate control groups begs the question of whether the provocative results presented are due to problems inherent to speech interfaces, as the author asserts, or whether the results are those to be expected in any multi-task environment.

In this session, paper #32 by Ponomarenko and Turzin, which was to discuss the effects of various environmental stresses which affect the implementation of automatic speech recognition technology in aviation was not presented.

6.0 CONCLUSIONS and RECOMMENDATIONS

The Symposium program offered an excellent sampling of relevant research from basic and applied laboratory studies to field and flight demonstrations of audio technologies. Many of the papers presented offered valuable reviews of previous work, as well as new data for the scientific community to consider. Particularly valuable were the overviews that were provided by selected technical experts for each of the three main technology areas. These overviews provided a context within which the audience could place the subsequent technical papers that were presented in each of the three areas. Among the conclusions to be taken away from this Symposium is the one that spatial auditory information promises to provide significant benefits in terms of target acquisition, threat avoidance, communications enhancement and situational awareness to the air crew of the future. At the same time it should be recognized for these benefits to be fully realized research is needed on auditory symbology and auditory icons. To efficiently convey information to the operator, 3-D audio systems have to be integrated into the communication infra-structure of the aircraft, high fidelity, wide bandwidth intercom systems with binaural output, and faster, more accurate head tracker technology need to be made available.

Automatic Speech Recognition technology also requires the development of the appropriate aircraft infrastructure, i.e., a high-fidelity, wide bandwidth intercom, before it is likely to become operational in aircraft.

Another conclusion that can be reached is that Active Noise Reduction technology has sufficiently matured; that its application in operational environments is underway. Despite that fact, work needs to continue on the development of advanced ANR technologies to enhance the noise attenuation available and to increase the frequency bandwidth over which the technology is effective.

One area that was covered by a single paper in this Symposium (paper #29) was the question of measures of communication effectiveness. Traditional measures of communication performance are either standardized intelligibility tests that provide a percent correct measure of word intelligibility without providing the answer to the question of what percent intelligibility is required in order to complete successfully a task within a specific operational context, or tasks designed ad hoc that have face validity for the communication task of interest, but are non-standardized and provide little, if any, statistical validity or reliability and are not

generalizable to other communication tasks. More research is needed for the development of reliable and valid measures of communication effectiveness in operational environments.

A second area that was only lightly touched upon in this Symposium, but which deserves further attention is the question of intelligibility of speech in noise for non-native speakers of the language. In the aviation context, this usually means that the language being used to communicate is English, while for one, or perhaps all of the communicators, English is not their native language.

This is an important question, particularly for a multi-national organization such as NATO. At this Symposium, only two papers mentioned this problem (the overview paper in the Speech Technology Session and paper #25) and in neither case was data on this question reported.

The available literature, while not extensive, does indicate that communications in which the language employed is not the native language of the listener require a higher signal-to-noise ratio than when it is the listener's native language. Even less is known about the problems that exist when the language is not the native language of the speaker. It is to be hoped that any future NATO symposia on audio technologies and communications will address these important human factors questions.

This is the second Symposium that the AMP has held addressing the use of the auditory channel in military operations, with emphasis on aviation. In 1981, a successful Symposium was held in Soesterberg, Netherlands, whose topic was Aural Communications in Aviation (Conference Proceedings AGARD-CP-311). At that meeting, one paper was presented on the topic of Active Noise Reduction. As we saw at the present conference, Active Noise Reduction is now being fielded with the operational forces. It is recommended that progress in the auditory sciences and technologies continue to be monitored by AGARD, and that when warranted, another symposium be sponsored to update the AGARD community on developments in the field with potential military applications.

Auditory Research in Denmark

S. Vesterhauge, MD, DMSc
Consultant, Senior Lecturer, Specialist Adviser
Dept. of Otolaryngology
Rigshospitalet, #2074
DK-2100 Copenhagen Ø
Denmark

If you ask an audiologist i Denmark when Danish audiology was born, you may have the answer, that it was born with the legislation of January 1950 which founded the Danish so-called Hearing Centres. Indeed, that was an important turning point, but it's not right. Danish auditory research began with the foundation of modern electronic science one day in 1819 or 1820, when the Danish physicist *Hans Christian Ørsted* (1777-1851), during a student lecture discovered that an electric current was able to deflect a compass needle. If nobody ever had discovered that phenomenon since then, where would we have been today in auditory research?

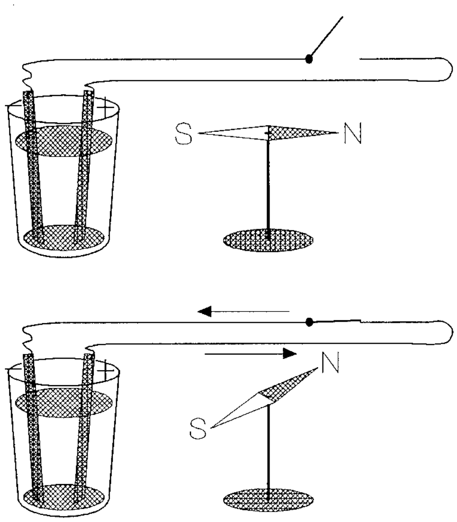


Figure 1. Hans Christian Ørsted's experiment.

Now, the audiologist wakes up - and I can hear his faint voice: *Don't you know that the first school for the deafs was established in Denmark twenty years before that* (it was in the town of Kiel in the principality of Holstein, which at that time was part of the Kingdom). In 1807 this idea was adopted in Copenhagen with the foundation of the Royal Danish Institute of Deafmutes. In this institution, the first physician of the institute, *P.A. Castberg* (1779-1823), some years later made unsuccessful experiments with galvanic stimulation of deaf children, and in the same institution, the Rev. Mr. *Rasmus Malling Hansen* (1835-90) in 1869 constructed one of the first typewriters in the world, the so-called writing ball, which became universally accepted and used in the years to come. So, we the Danes were there from the very beginning - no doubt.

The hearing centres founded by the law of 1950 were welfare institutions, where deaf people could be examined, diagnosed and treated with hearing aids.

The old problem of whatever came first, the egg or the hen, can be solved in this case, if the centres are con Danish electronic industry, a small number of brilliant Danish ENT specialists, interested and experienced in hearing research and an old Danish tradition of education of deaf children - so to say were the egg from which the hearing centres were hatched.

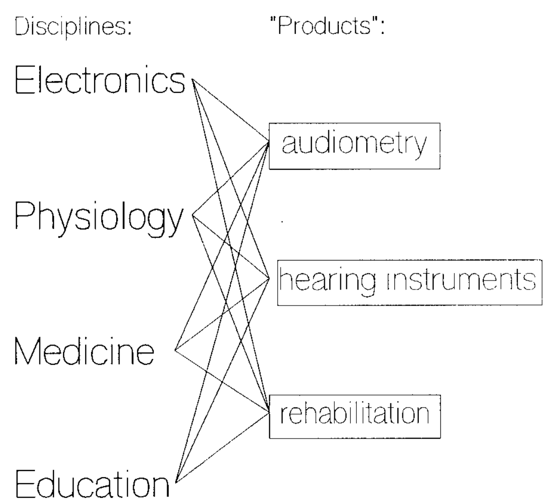


Figure 2. The network of Danish research disciplines and their "products".

The electronic industry was represented by the three Danish hearing instrument manufacturers, *Oticon*, *Widex* and *Danavox* and by the well known institution *Brüel & Kjær*, renowned for its sound level-meters and condensator microphones. Further a Danish audiometer manufacturer, *Madsen Electronics* was part of scenario. Fortunately this particular hen, the Act of Welfare of the Hearing-Impaired, was fertile enough to lay a large number of precious eggs before it was more or less sacrificed by the Danish Social Security Act of 1976.

The hearing centres were headed by ENT specialists - that's one of the reasons why in Denmark audiologists are physicians by education. A number of these pioneer audiologists maintained and developed the scientific network which included electronic engineers, the devel-

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

oping Danish electronic industry, and teachers and psychologists, which all together are the basis of the present and past Danish auditory research. I will try to provide you a survey of this field of research in Denmark, emphasizing a number of Danish pioneer scientists and research products.

Otto Metz (1905-93) was the real father of Danish auditory research, he was an ingenious ENT surgeon, who during World War II developed methods to describe and measure the acoustic impedance of the human ear. Metz had a Jewish background and fortunately he escaped to Sweden during the German occupation and returned in 1945 to complete his pioneer works in the years to come.

Metz's efforts were crowned by the works of *Otto Jepsen* and *K.A. Thomsen* developing and describing methods to clinically test and measure the stapedial reflexes and the middle ear pressure.

Otto Metz:	1905-93. Thesis 1946: The acoustic Impedance Measured on Normal and Pathologic Ears.
Otto Jepsen:	1916- . Thesis 1955: Studies on the Acoustic Stapedius Reflex in Man.
K.A. Thomsen:	1920- . Thesis 1958: Akustisk impedansmåling ved funktionsundersøgelser af tuba Eustachii og til bestemmelse af recruitment.
K. Terkildsen:	1918-84. Thesis 1963: Akustiske impedansmålinger og mellemørets funktion

Terkildsen K, Scott Nielsen S. An Electroacoustic Impedance measuring bridge for clinical use. Arch Otolaryng 1960;72:339

Figure 3. Pioneering Danish scientists and their principal publications.

K.A. Thomsen who made the pioneer work in middle ear pressure measurement, tympanometry, using the impedance methods developed by Metz. Both Jepsen and Thomsen are still going strong. I met and talked with both of them just before this meeting of our national scientific otolaryngological society.

The techniques developed by the audiologist *Knud Terkildsen* and the engineer *Scott Nielsen* in 1960 announced new perspectives in clinical audiometry and managed to conquer and establish a prominent position in our clinical daily life during the sixties and the seventies. They developed the so-called electro-acoustic bridge. This old photo shows Terkildsen testing Scott Nielsen by means of their impedance audiometer. Scott Nielsen is still doing fine, but unfortunately we lost dr. Terkilsen twelve years ago. He was the generator behind so many scientific projects of our department including my own thesis.

Please, let me describe shortly what impedance audiometry is to those who are not familiar with the technique.

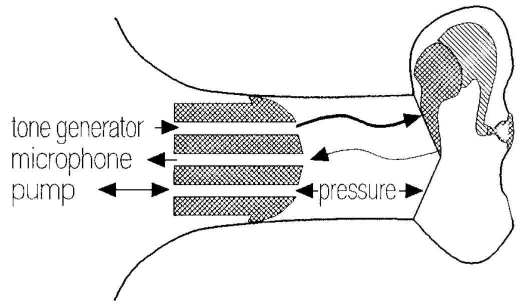


Figure 4. The priciples of middle ear impedance measurement.

A small three canals probe is inserted into the ear canal. Through one of the canals, a 220 Hz tone is transmitted to the ear. Most of the energy of this signal is transmitted through the tympanic membrane and the ossicular chain to the inner ear. Some of the energy is reflected from the ear drum and the relationship between input and output sound energy can be validated by means of the electro-acoustic bridge. The impedance (or more precisely, the immittance) can be altered by two different means during the test. A contraction of the stapedial muscles caused by a high sound pressure in either ear will result in a higher tension of the ear drum resulting in the reflection of a larger part of the energy presented to ear and the transmission of a lesser part of the energy to the inner ear. This phenomenon can be recorded by the impedance audiometer and thresholds of stapedial reflexes, which are of great clinical value can be measured. This was the method developed by Otto Jepsen.

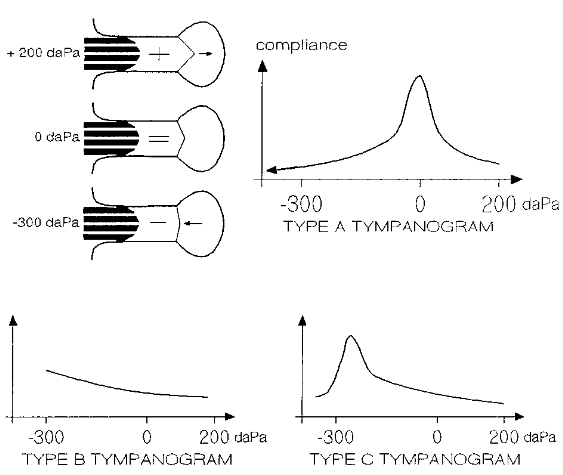


Figure 5. The priciples of impedance tympanometry. For explanation, see next page.

By changing the air pressure of the closed compartment between the probe and the tympanic membrane the amount of sound energy reflected from the ear drum can be manipulated, see fig. 5. When the pressure on the two sides of the ear drum is equal, the amount of energy reflected from the ear drum is minimal. By means of pressure variations in the ear canal, the pressure of the middle ear can be estimated quite precisely. The type A tympanogram in fig. 5 displays the impedance variations in middle ear with a normal pressure when exposed pressure variation in the ear canal. The type C and B tympanograms appear when pressure is low and the middle ear is fluid filled, respectively.

This is tympanometry as done all over the world due to the pioneering work of Metz, Thomsen, Terkildsen and Scott Nielsen. The Danish company collaborating with this group of scientists, Madsen Electronics, was the first to introduce equipment usable for this purpose on the world market.

Another area of manufacture and research in the auditory field in Denmark is hearing instruments. The first progress in Danish hearing instrument manufacture came when electronic miniaturation became actual in the fifties. The manufacturers utilized the small components and made eyeglass instruments, hair ornament and behind-the-ear instruments. In the seventies, equipment to measure actual hearing instrument input to the patients' ears by inserting a small tube connected to a microphone into the ear canal between the earmold and the ear drum was produced. This resulted in much more accurate adjustment of hearing instruments - a large benefit for the patient. Further they demonstrated that individual earmold design had considerable influence on the performance of the instrument.

Recently two of the three Danish hearing instrument manufacturers have impressed the whole world by introducing fully digitized hearing instruments to the market. They were the first to do so, and hopefully they will be able to maintain the Danish share of the world hearing instrument market, which uses to be approximately 25%. It's really impressing that you can fit a Pentium processor into the ear canal. The real benefit of this type of hearing instruments is that the amplification of the instrument now can be adjusted exactly according to the special needs of the patient.

Last but not least. As a direct continuation of the tradition which began with dr. Metz's work concerning ear impedance, which after 15 years resulted in the development of Terkildsen and Scott Nielsen electro-acoustic impedance bridge, the engineers of our department have been able to construct a prototype of a set-up that makes it possible to exploit the fact that the ear actually emits sound, the so-called otoacoustic emissions. The equipment is used make sure that newborn babies are able to hear. It analyzes the otoacoutic emission responses produced by short acoustic stimuli presented to the ear in a number of thousands within a few seconds.

References:

1. Bergenstoff, H. Hearing Instruments from Past to Present, *in* Recent Developments in Hearing Instument Technology, 16th Danavox Symposium, eds. Joel Beilin & Gert R. Jensen, Copenhagen 1993, pp 13-38.

The Audio Environment in Aircraft

Dr Graham Rood
Research Manager Man-Machine Integration
Systems Integration Department
DRA Farnborough, Room 204, Bldg R177
Hants, GU14 6TD, U.K.

Above 10,000 ft the DH4 could outfly contemporary single-seat fighters, but, if caught, was usually an easy victim because the cockpits were so far apart that the crew was forced to rely on the Gosport Tube as a means of co-ordinating their defence in the noise and heat of battle; this proved virtually useless.

from "De Havilland Aircraft" by A.J.Jackson

'There is no homecoming for the man who draws near
to them unawares and hears the sirens voices
to prevent any of your crew from hearing, soften some
beeswax and plug their ears with it.'

- Homer, *The Odyssey* (Translated by E V Rieu)
from 1979 paper by Edgar Shaw NRC Canada

1. INTRODUCTION

Historically, noise levels in cockpits have always been high, and even in biplanes communications were sometimes a problem - even a DH4, and it was the positioning of pilots in open cockpits between the engines of the Handley Page passenger aircraft type, in post WW1 commercial aviation, and the long haul flights with constant exposure to engine noise, that further highlighted the issue of hearing loss and the 'Aviators Notch'. This has continued into the monoplane era, and during WW2, when noise became a particular problem and reliable measurements were made, Ref 1, noise levels were high in many combat aircraft, mainly from engine, exhaust and propeller noise - although not exclusively. High levels of noise could be generated from other aircraft/engine systems, and an example is shown in Fig 1 of cabin noise from a conventionally powered, unsupercharged, twin engined Dornier 17 - with the highest noise levels at the propeller frequencies and in the propeller line - and a Junkers 87 Stuka - single engined - where the cabin noise is dominated, at the higher frequencies, by supercharger noise, Fig 2, Ref 2.

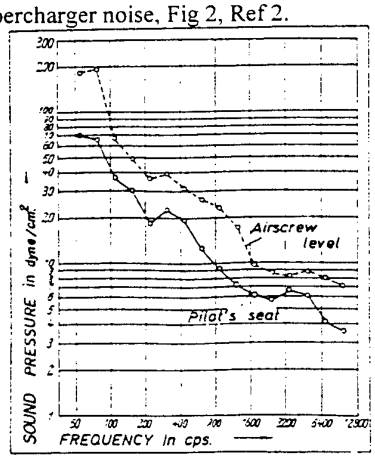


Fig 1: Cabin Noise in Do17 Aircraft

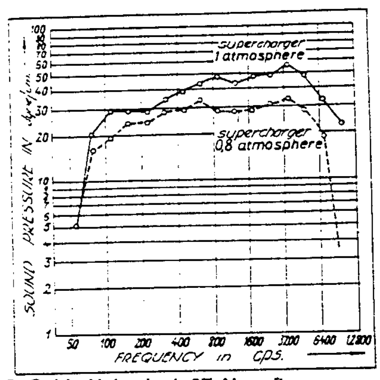


Fig 2: Cabin Noise in Ju87 Aircraft

Figs 3 & 4 illustrate the levels of noise exposure in a number of other WW2 USAAF aircraft from Ref 1 and illustrate the high levels of exposure, generally in the 120 dB OASPL region. The progression to the gas turbine engine removed the propeller and exhaust noise and cockpit noise levels were reduced, Fig 5, Ref 1, and the gradual movement of the engine(s) towards the rear of the aircraft or buried in the fuselage, further helped the acoustic environment.

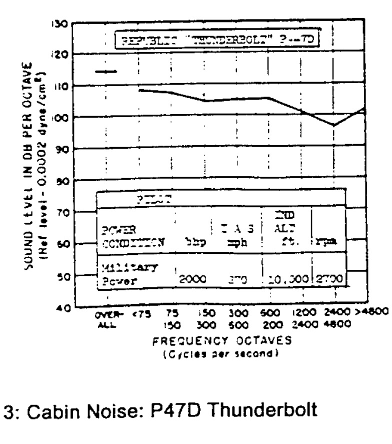


Fig 3: Cabin Noise: P47D Thunderbolt

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

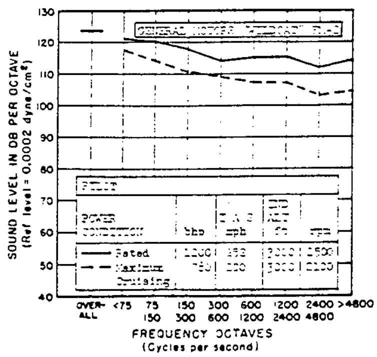


Fig 4: Cabin Noise. Wildcat FM2

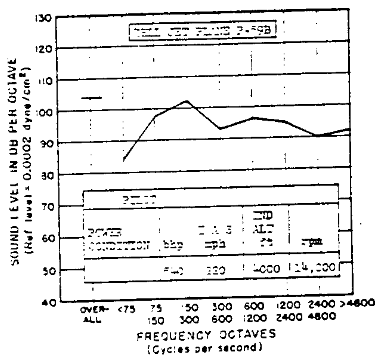


Fig 5: Cabin Noise Bell P59B Aircobra

There are two main categories of aircraft noise; that perceived externally, mainly on the ground and that perceived internally in the aircraft cockpit or cabin. Both are worthy of analysis and discussion, but, for this paper, only the internal noise generation of the cockpit & cabin noise is discussed, and this is based upon the, not unreasonable, premise that cabin noise is the greatest contributor to the interference with audio communications.

2. CURRENT STATE

The move towards gas-turbine engines and higher aircraft speeds, however, generated a new series of noise problems.

The majority of current problems from high levels of internal cockpit noise, arise, essentially, from the post 1960's need to fly operationally at high speed and low-level as part of tactical flight, in order to minimise detection by radars and other sensors and minimise exposure times to ground based weapon systems. The ingress to target is usually flown at speeds around 420 to 480 knots at heights at or below 250 ft and egress is quite often lower and faster. At these speeds and heights noise levels in the high speed jet aircraft cockpit, or cabin, have been increasing over the years, with one or two exceptions, and Fig 6 illustrates this trend for UK aircraft.

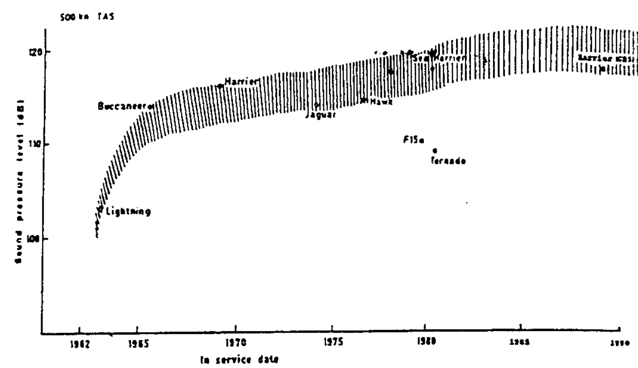


Fig 6: Historic Trend of Cabin Noise Levels in UK Fast Jet Aircraft

Cabin noise levels of 115 dB to 120 dB SPL are not unusual in high-speed low-level flight, and a comparison of three aircraft, varying from "noisy" to "acceptable' are shown in Fig 7. Similar figures can be shown for most aircraft, independent of the country of manufacture, and the cockpit noise figures will generally be a direct function of the operational requirements for external visual fields (largish bulbous canopy) and aspects such as the type of escape (canopy thickness) in the cockpit design.

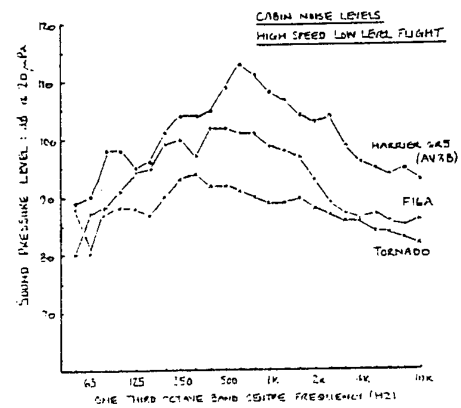


Fig 7: Cabin Noise, High Speed Low-level

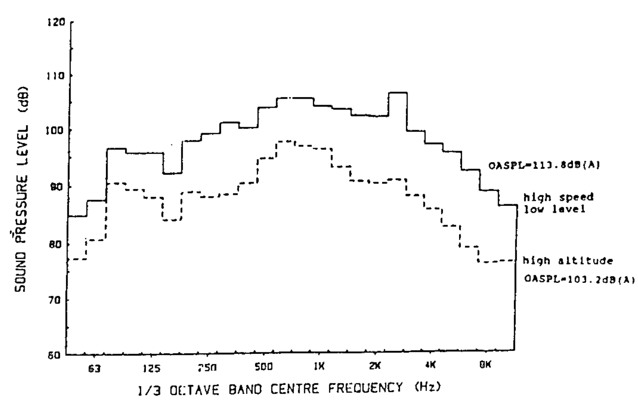


Fig 8: Effect of Altitude on Cabin Noise

In these high speed jets, the cabin noise spectrum is generally random in nature and broad band, and the noise is generated from two predominant sources. One is from the external airflow around the aircraft canopy and the front structure of the aircraft and the other is from internally generated noise from the air conditioning and cooling flows into the cockpit space. Generally the noise levels generated from the external airflow sources are dependent upon the dynamic pressures on the aircraft ($\frac{1}{2}\rho v^2$) and thus speed and height, and Fig 8 illustrates the change in cabin noise with altitude. Manoeuvres that further alter the instability of the flow patterns around the canopy and aircraft front fuselage will also increase noise levels. Cabin conditioning/cooling flow noise levels are nominally constant with speed & height, but with some changes in noise spectrum due to conditioning mode, Fig 9.

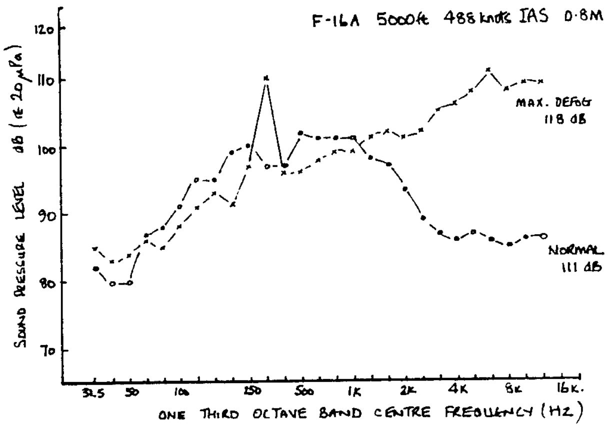


Fig 9: Effects of Cabin Conditioning Modes

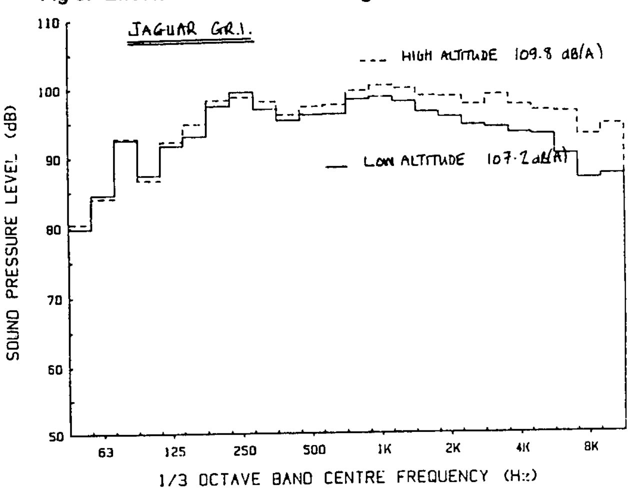


Fig 10: B.Ae. Jaguar Cabin Noise

Thus, depending upon the design of the aircraft and its systems, the cabin noise may be dominated by either the externally generated, or internally generated, noise or be a balance of both of these noise sources. The Jaguar is an example where the contributions from both sources are approximately equal, and the cabin noise remains essentially constant irrespective of speed or height, Fig 10.

However, in some aircraft, there are other contributing factors. In the British Aerospace Harrier/AV8B, for instance, there is a contribution from the engine compressor fan, Fig 11. Aircraft of the Harrier type, which have the ability to hover, need a large compressor fan to meet the engine airflow requirements with no forward speed and thus have a large fan close to the cockpit, and this is seen as a discrete narrow band noise source in the 2.5 kHz area, and obviously dependent upon engine speed.

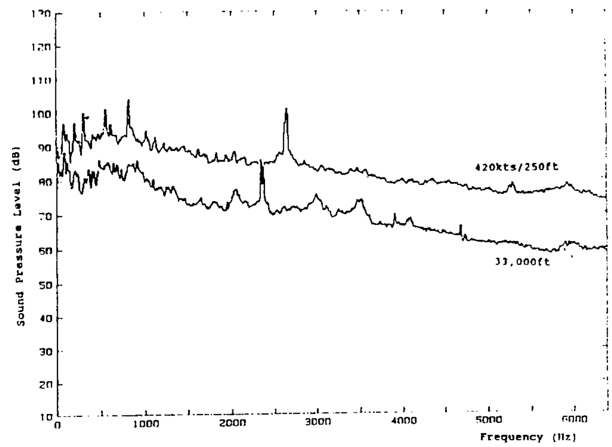


Fig 11: B.Ae. Harrier GR5 Cabin Noise Levels; Narrow Band & One-third Octave Band Analysis

Helicopters have a different mechanism of generating noise, and the predominant helicopter noise is generated in narrow band discrete tones and associated harmonics. The sources are generally both aerodynamic and mechanical. Aerodynamically induced noise is generated from the main and tail rotors, any interaction between the rotors in a twin rotor design (e.g. Chinook) or between the rotors and fuselage; and the mechanical noise originates from revolving systems connected to the rotors in the form of gearboxes, transmission shafts, transfer gears, auxilliary systems drive shafts etc. Fig 12 shows narrow band analyses for two helicopters and the sources of the noise peaks. Due to each type of helicopter being mechanically and aerodynamically different (e.g. 2, 3, 4, 5 or more rotor blades in the main rotor, or differing gearing ratios in the main gearbox etc.), each helicopter will have a unique acoustic signature. Boundary layer noise is not present to any great extent due to the restricted forward speeds of helicopters, but turbulent airflow noise will be apparent when the helicopters are flown with doors, windows or ramps open. Some helicopters (e.g. OH58D etc.) have significant amounts of noise generated from avionic systems equipment installed in the aircraft and cooling fans and other noise generators in this equipment may add significantly to the overall cockpit/cabin noise levels.

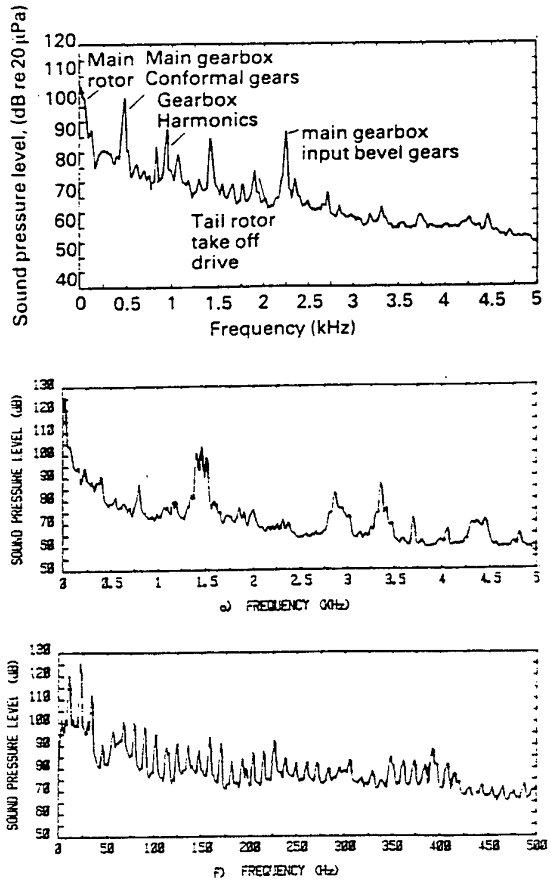


Fig 12: Chinook and Lynx Cabin Noise Analyses

Aircraft that either fall between being a helicopter or a fast jet (i.e. transport aircraft of the Hercules type (with propellers), Fig 13, Ref 3, or C17 type (with gas turbines)) or use the Tilt Rotor approach, have some noise generated from propellers/rotors or wing mounted gas turbines, some from boundary layer flow noise, often over the wing slots and slats that assist in the lift process, and some from equipment cooling and/or cockpit conditioning systems, and thus are a differing combination of discrete and random noise.

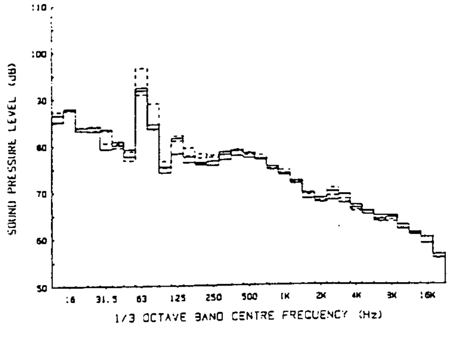


Fig 13a: RAF Hercules C1/3 Cabin Noise Analyses

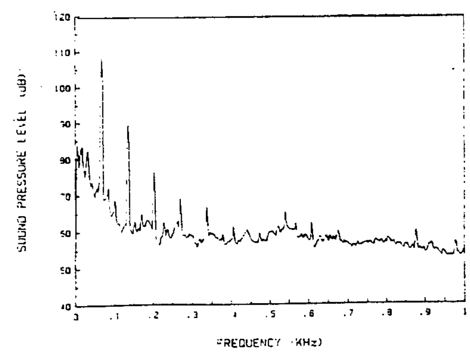


Fig 13b: RAF Hercules C1/3 Cabin Noise Analyses

Aircraft that are essentially civil-based militiarised aircraft - surveillance / maritime patrol of the Nimrod (DH Comet) type or Command & Control/AWACS/JSTARS (Boeing 707) type generate a small amount of boundary layer noise, mainly in the front cockpit,, but the predominant source is from the forced airflow (random cooling-duct outlet noise and associated discrete fan noise) in the aircraft to cool the avionics and crew in the rear cabins of the aircraft

3. EFFECTS OF NOISE

Whilst high levels of noise can have a wide range of effects, in the aircraft cockpit or cabin the effects can be generally restricted, in military terms, to two main areas:

1. Hearing Damage Risk, and
2. Interference with Communications & Listening Tasks

3.1 Hearing Damage Risk

In terms of hearing damage risk, the two main contributors are noise levels at the ear from the cabin noise and the contributions made by the acoustic levels of the speech communication and signal communication levels. Even with the protective helmet, cabin noise levels alone at the ear can be high enough to produce a risk of hearing damage. On top of the risk generated from cabin noise levels will come the additional contribution from the communications and in operational measurements in helicopters (Ref 4) and fixed wing operations (Ref 5) figures in the region of an average of 6 to 10 dB(A) can be added to that of the noise levels alone, to attribute the additional contribution from the communication levels.

The current European recommended limit is 85 dB(A) for an 8 hour daily exposure, and, for most military forces, this is a target figure - not necessarily mandatory. With this 85 dB(A) figure, it is possible to trade the noise level against the time of exposure and if the noise levels are lower by one half (i.e. a reduction of 3dB(A)) then exposure times can be doubled (to 16 hrs). However, if the levels are increased by 3 dB(A) to 88 dB(A) (i.e. doubled) then the exposure times must be proportionately decreased (i.e. halved) to 4 hrs. Thus, following this logic, exposures of 100 dB(A) should only be tolerated for 15 mins per 8 hour day, and so on.

Measurements of hearing damage risk in the RAF fleet over a number of years has shown improvements in the lessening of the risk, predominantly from the use of the Mk4 and Mk 10 flying helmets with better acoustic attenuation than the previously used Mk 2/3 series, but total noise exposure figures, taken at the pilots' ear, still have time corrected L_{eq} values in the region of 85 to 92 dB(A), Ref 6.

3.2 Interference with Communications

Such high noise levels in the aircraft cockpit or cabin, and the consequent high noise levels at the aircrew ear and lips, generate interference with speech, and non-speech, communications.

3.2.1 Speech Communications

In speech communications, noise is introduced into the communications line through the microphone. In a helicopter, the microphone is fully exposed to the cabin noise, and helicopter microphones are generally of the noise cancelling type, where **some (not all)** of the noise is cancelled. Generally this is at frequencies below 1kHz and depends upon the physical design characteristics of the microphone. Due to the need for robustness in microphone construction for the aircraft environment, most cancellation is below 500Hz - which is more appropriate for the helicopter environment with its predominance of low frequency noise.

In the fixed wing cabin, aircrew generally wear an oxygen mask which incorporates a microphone. Noise enters the mask, and thus into the speech line, in a number of ways. The passive attenuation characteristic of a mask, Fig 14, Ref 7, is similar to helmet attenuation, and thus the interior of the mask is rich in low frequency noise. Whilst breathing out, the expiratory valve in the mask opens and generates a direct path to the outside noise, and to the microphone. The oxygen hose is exposed to the full cockpit noise and thus the noise transmitted through the hose walls to the hose interior can be picked up by the mask microphone, and further noise can be generated directly from the oxygen supply system. These are the mechanisms of the generation of noise in an oxygen mask and the contributions of each area to the overall levels will vary with the particular equipments and cabin noise levels.

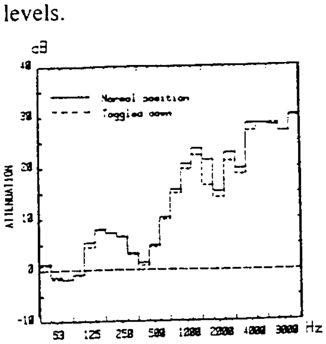


Fig 14: UK Oxygen Mask (P/Q type) Attenuation

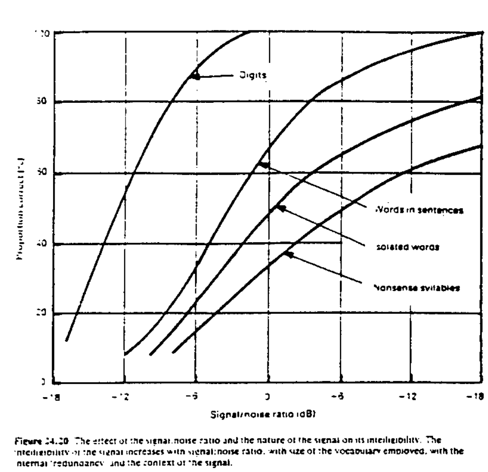


Fig 15: Speech Intelligibility Curves

The mask microphone frequency response characteristic is adjusted by taking into account both the free field speech levels and the effect of enhancing the low frequency content of the speech by talking into a small (by acoustic wavelength standards) cavity mask, and the free field frequency response of the microphone. The overall frequency response of the mask/microphone system is tailored to give as flat a response as necessary. Noise in the mask is apparent from the mechanisms outlined above, and is generally predominantly cabin noise, and this is added to the speech signal at the microphone. The ratio of speech to noise is called, not unsuprisingly, the speech signal (S) to noise (N) ratio and is expressed in dB of speech above the noise (e.g. 12 dB S/N ratio).

As the S/N ratio increases there is a corresponding increase in speech intelligibility and a plot of S/N ratio against speech intelligibility, Fig 15, shows a characteristic 'S' curve. Thus, as signal to noise ratios increase, speech intelligibility increases until an asymptotic point is reached. Above a given signal to noise ratio improvements in speech intelligibility are marginal.

There is however, a further factor in intelligibility, and that is involved with the contextural information within the speech. If the text has a measure of redundancy, then some losses in intelligibility can be recovered by the human from the overall contextural meaning. If, however, there is no redundancy in the message, as when nonsense words are used, then the speech signal to noise ratios must be correspondingly higher. Fig 15 shows clearly this effect.

Whilst the human brain has an ability to tease out the signal from the noise, using the differing characteristics of speech and noise, and even at a S/N ratio of 0 dB will reliably understand some communications (48 % for isolated words in Fig 15), machines recognising speech may not always be as

immediately skilled. Some HF secure speech systems which use an encoder and decoder of the LPC10 type, at least in NATO, and rely upon the decomposition of the microphone signal (encode) and the subsequent reconstitution (decode) after transmission, have problems with high noise environments. This is generally caused by the encoding of the pressure signal from the microphone, which is a combination of noise & speech, but which is not differentiated by the encoder. Subsequent reconstitution of the encoded speech attempts to make voicing sounds out of the overall pressure signal, which introduces varying levels of distortion, and thus unintelligibility, into the reconstituted signal. Signal processing of the microphone signal to reduce the noise may improve intelligibility and the use of contextural information in the messages also can improve the overall intelligibility. However, with the gradual move to lower bit rates in speech communication systems, efforts will have to be made in speech pre-processing, or alternative forms of low bit-rate transmission found, in order to maintain operationally acceptable levels of speech intelligibility in noisy environments.

Speech signal to noise ratios can either be degraded at the speaking end of the communications chain (i.e. the microphone) or the listening end (the pilots ear), or, in many cases, both ends! While signal processing may take partial care at the microphone end, at the listening end the pilot can essentially only increase the speech intelligibility by increasing the signal to noise ratio by turning up the communication level gain (if there is any left).This, unfortunately, increases the risk of hearing damage, as the contribution to the overall damage levels from the communications increase. A better solution is to improve the S/N at the ear by decreasing the noise levels, and this can be accomplished either by good passive attenuation alone or by a combination of active and passive using Active Noise Reduction techniques. Recent laboratory trials in helicopter noise (Ref 8) have shown increases in intelligibility of some 7%, and more where noise levels are more intrusive. In low noise environments the use of ANR is generally unnecessary, as signal to noise ratios are adequate for high intelligibility, but, as cabin noise levels become more intrusive, the extended use of active reduction techniques to supplement the, essentially, static growth of passive attenuation, will become progressively more appropriate and necessary.

3.2.2 Speech communications to Machines

Other classes or types of machines are affected to various levels by cabin noise and speech recognition devices or Direct Voice Input machines accuracies are reduced in the high noise environment. The speech recognition accuracy, however, can be improved by suitable training templates and accuracies are in the region of 96% in normal flight conditions (Ref 9).

3.2.3 Unaided Speech Communications

The previous paragraphs have discussed speech intelligibility where communications is aided by an electronically amplified communications system. Under some conditions (e.g. where a commander is briefing and updating his troops in flight in a helicopter en-route to the battlefield), it is not always possible to used aided communications, and unaided communications are appropriate. Under these conditions a process called Speech Interference Level (SIL) or Preferred Speech Interference Level (PSIL) can be used to determine whether direct speech communications can be effective in a given noise environment. A series of experimental trials (Ref 10) have resulted in a set of tables, which, when the noise environment is known, specify the attainable unaided communication as a function of speaking level at speaker separation (i.e. normal, raised, shouting etc. speech levels at 0.5, 1 or 2 meters or feet. etc.).

As well as for previously mentioned use in helicopters, this type of analysis is also appropriate for non-aided communications in transport or civil cockpits, or in radar or comms operators positions in aircraft of the E3D/AWACS/JSTARS type.

3.2.4 Non Speech Communications

Non-speech signals, such as Auditory Warnings, Navigation signals, Weapons status signals etc., can also be masked or degraded by noise. Over the last ten years, considerable research has been undertaken (Ref 11 to 13) to set rules and methods for the design of Auditory Warnings or Auditory Icons in high noise conditions. With the knowledge of the noise levels at the pilots' ear, both the spectral content and sound pressure levels of the warning sound or icon can be calculated that will allow a 100% chance of detection by the pilot, without having to be exposed to excessively loud, and unacceptable, levels at the ear. Figs 16 & 17 illustrate the approach and shows some examples of a comparison of acceptable design levels against some signals in operational service (Ref 14).

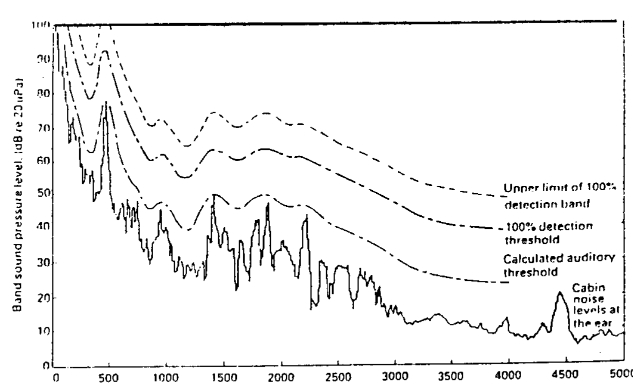


Fig 16: Auditory Warning Acoustic Level Principles

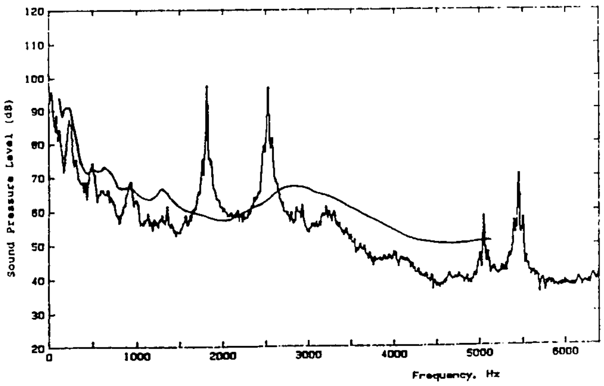


Fig 17: Measurements of Warnings in Flight

4. PROTECTION

Since noise is generated by the flow disturbance around the canopy and fuselage, then noise levels may be altered by changes in shape of either the canopy or fuselage. In practice this can generate more severe aircraft performance penalties (e.g. aerodynamic drag, aircraft stability etc.) and whilst it is theoretically possible and has been implemented in the past in civil aircraft (Fig 18), in most cases, for military aircraft, such changes are not practically possible on noise reduction grounds alone.

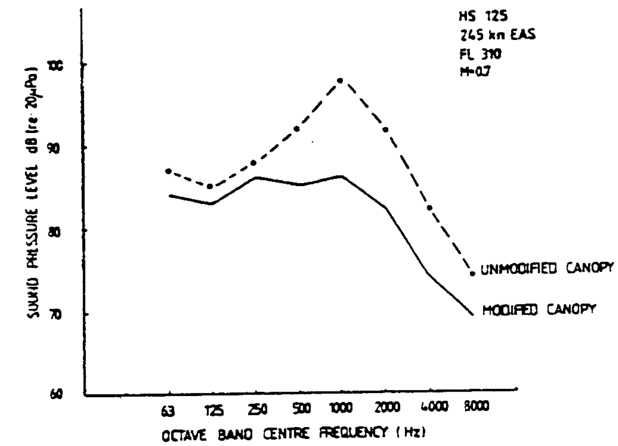


Fig 18: Effects of Cabin Shape on Internal Noise Levels

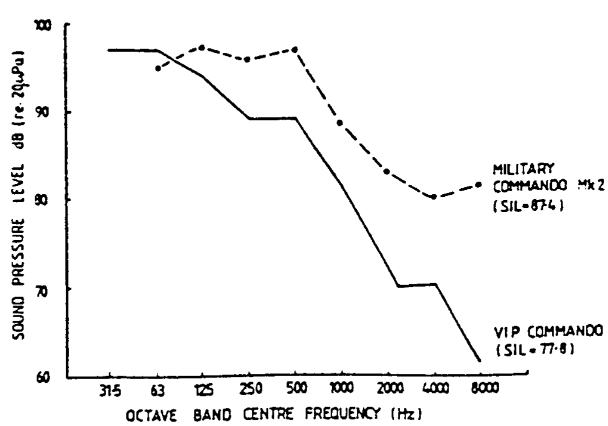


Fig 19: Effects of Soundproofing in a Helicopter

An alternative approach is to use soundproofing, and this is regularly used in civil airliners and in some special cases of military aircraft - Fig 19 shows the reduction of cabin noise for a VIP helicopter compared to a similar military airframe. Whilst demonstrably effective, soundproofing occupies volume and adds mass, and these two solutions, which add perceived non-operational weight and reduce the payload volume, are not generally compatible with military operational ideas. In practice, soundproofing has to be maintained intact to be fully effective, and the military operations, particularly in helicopters, makes it difficult to maintain, although wide use is made of such noise reduction techniques in helicopters.

With these practical limitations, the most effective solution, both in terms of cost and operational effectiveness, is to provide individual protection on the pilot/aircrew.

Within most military cockpits, aircrew are required to wear a protective flying helmet, and this helmet can be made to provide a level of acoustic protection, generally by incorporating circumaural hearing protector shells into the helmet. The shells provide an overall protection, but the levels of protection change with frequency. Circumaural protectors generally have three mechanisms of protection, each in a particular frequency band.

1. Up to frequencies in the region of 300 to 400 Hz, the noise levels at the ear are controlled by the volume of the earshell and the stiffness of the acoustic seal. As the shell moves on the spring stiffness of the seal (and the human flesh), the changing volumes create a corresponding pressure change and this limits the acoustic attenuation at these low frequencies. If a stiffer seal is used, say a liquid seal, then the increased stiffness limits the shell movement and the consequent pressure changes - resulting in more attenuation.

2. Above this frequency and up to about 2 kHz, the attenuation is controlled by the characteristics of the materials used in the construction of the shell and the internal damping properties of the material. The attenuation will then generally follow a 12 dB/octave slope.

3. Above 2 kHz, control of the attenuation is from damping and absorption of the higher frequency resonances that occur at these shorter wavelengths by the use of foam or fibrous based materials put into the shall cavity to effect the damping and absorption of the sound waves.

The overall effect of these mechanisms is to produce an attenuation characteristic shown in Fig 20, and this is a characteristic that is typical for all types of

circumaural hearing protector - military or civil, flying helmet mounted or headset mounted for military or industrial factory floor use.

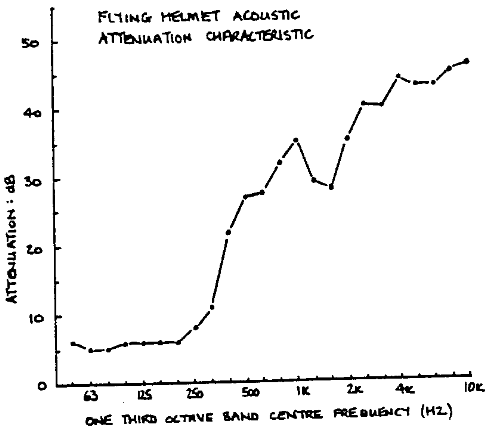


Fig 20: Characteristic Attenuation of a Flying Helmet

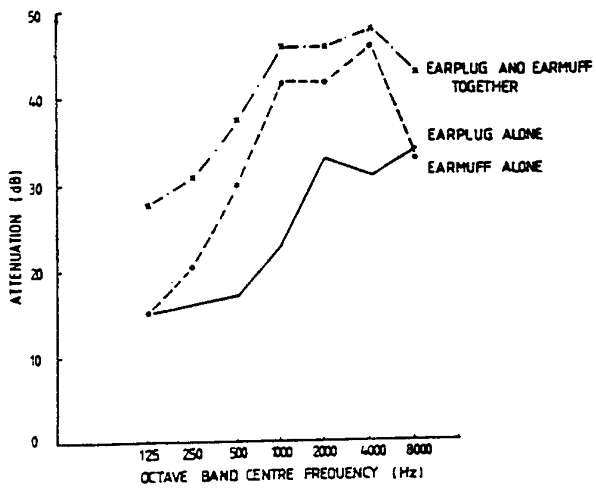


Fig 21: Noise Attenuation of Earplugs & Earmuffs

As is the case for most engineering systems, some protectors are better than others, some companies understand the design process better than others and some sacrifice good design & performance for lower cost.

Changes can be made to the attenuation characteristic by the use of different materials (in the mid- frequency range), different internal absorbent materials (at higher frequencies) or by the increase of shell volume (at low frequencies). Doubling the volume of the shell will provide a theoretical increase in low frequency attenuation of some 6dB and a further doubling will provide a further 6dB increase, and so on. However, practicality of use, particularly in the aircraft cockpit, precludes the use of the physical size of helmet that will accrue from these large shell sizes.

An alternative is to use earplugs whose mechanism to reduce noise is to occlude the ear canal. Like circumaural protectors, there are a number of types, all with differing levels of performance - but essentially there are the harder plastic type or foam plug type. Earplugs generally give better passive low frequency attenuation than circumaural devices, (Fig 21) but are subjected to the same performance limitation mechanisms as circumaural protectors. Some military forces allow the use of earplugs under flying helmets (sometimes under the circumaural protector) and the use of these two devices together will increase the overall attenuation marginally Fig 21. With earplugs in use inside a helmet, problems can arise from the occluding of the ear canal reducing not only the noise but also the communications from the helmet. Not all communication systems design will allow the volume of the communications to be increased to a level to compensate for the acoustic reduction of the ear plug - at least not without distortion of the communications signal. However, earplugs are now available with communications transducer inserts to alleviate that particular problem. Aeromedical problems may occur in fixed wing aircraft, in the form of differential pressure changes between the outer and inner ear during explosive or rapid decompression. This is not perceived as a particular problem in helicopter operations.

Because of the relatively poor attenuation at the lower frequencies, both from circumaural protectors and from earplugs, coupled with the high levels of cockpit noise at these frequencies, the noise levels at the pilots ear (Fig 22) are rich in low frequency content. Passive methods are available, in terms of large volume shells, but are impracticable, and the approach started some 20 years ago (Ref 15, 16) and discussed at the last AGARD conference on Acoustics (Ref 17) was to use active methods of cancelling the noise - Active Noise Reduction (ANR).

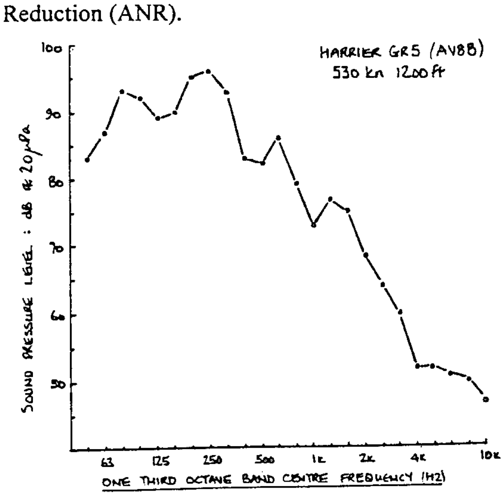
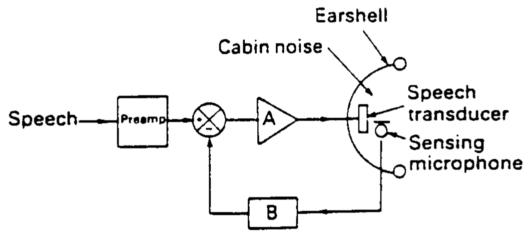


Fig 22: Noise at the ear of a Harrier GR5 Pilot



Schematic diagram of an active noise reduction system. The diagram shows how the noise inside the earshell is collected by the sensing microphone and fed back in a negative feedback loop via amplifier B, to be inverted in phase and fed through amplifier A back into the earshell via the speech transducer. The inverted phase noise from the feedback loop and the in-phase noise already in the earshell are mutually destructive and the sound pressure levels in the earshell are consequently reduced. Since speech is also reduced in level, it is preamplified from its source and fed into the shell at an increased level, thus compensating for the active reduction of the speech levels.

Fig 23: Principles of Active Noise Reduction Systems

The principle of Active Noise Reduction is relatively simple, but, of course, the practical application is somewhat more difficult.

In principle (Fig 23) the noise inside the earshell is sampled with a microphone, passed through inverting electronics, which changes the phase, and injected back into the same shell with this phase change and destructive interference of the noise in the shell occurs. A number of systems exist in the UK, USA, France, Netherlands etc. and a typical active attenuation performance is shown in Fig 24 Within an aircrew earshell, the active working range is between 50 Hz and 1000Hz (500Hz only for some systems) and peak levels of active attenuation are up to 20 to 23 dB. When added (arithmetically) to the existing passive attenuation of the shell Fig 25, significant overall attenuations are apparent, and in operational flight trials, Refs 18 & 19, and laboratory trials, reductions of around 14 dB(A) are possible in both fixed and rotary wing aircraft. Similar reductions are available from Active Ear Plugs - ANR incorporated into earplugs.

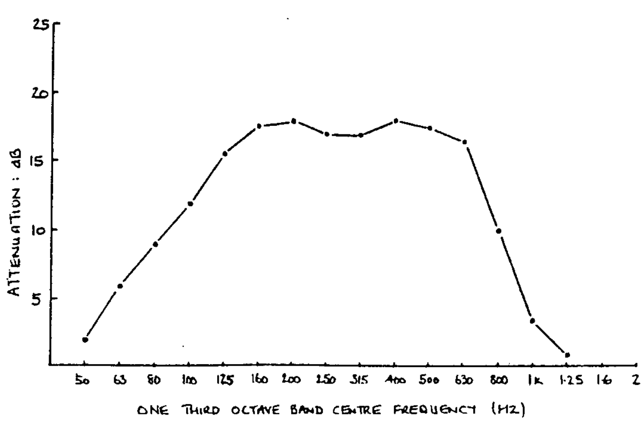


Fig 24: Typical Active Noise Reduction Spectrum

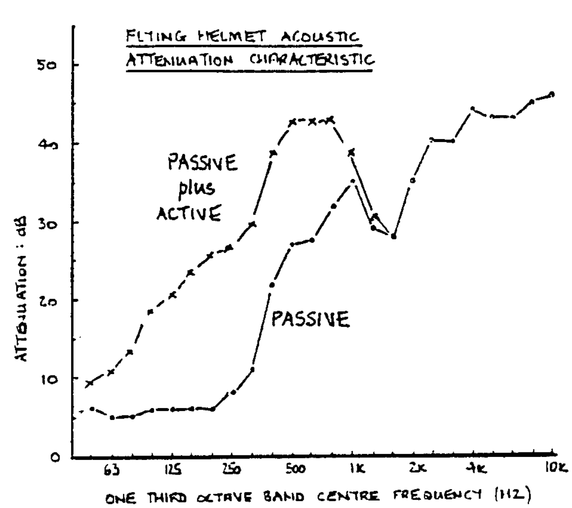


Fig 25: Overall Attenuation Characteristic (Active plus Passive)

In terms of reduction of hearing damage risk, even with a 10dB(A) reduction of noise at the ear, this is a significant reduction in risk. Alternatively, this level of reduction will allow aircrew to fly more than ten times the number of flying hours per year without any increase in risk of permanent hearing damage over that already experienced.

5. SUMMARY

In summary, the overall aim of much of the Acoustic & Noise research is to minimise the risk of hearing damage whilst maximising the operational communications capability, with communications meaning all necessary signals to the pilots' ear. Calculations show that by using next generation active noise reduction technology in the flying helmet (Ref 20), producing higher levels of active reduction or combinations of active & passive attenuation, it is possible to reduce the noise levels at the pilots ear to around 75 dB(A), such that the hearing damage risk is essentially reduced to zero. The reduction of noise levels at the ear is also fully compatible with improving speech & non-speech communications. At the talker & signal input end - at the microphone - signal processing approaches are needed to provide adequate signal to noise ratios for transmission, not only for the reception by humans, but also for the recognition by machines, whether they are part of a human centred system (e.g. Vocoder) or a machine centred system (e.g. Voice Recognition Systems). At the listening end, research into means of noise reduction, either active or passive - or, more likely, both - will support the overall aims.

The use of Auditory Displays to enhance operational effectiveness, both through the use Spatial Localisation of Sound (SLS) and the associated use of well designed

and tested Auditory Icons, will require the use of higher quality transducers in the helmet earshell, as will the use of good performance ANR, and this will support the move towards higher speech intelligibility. Overall, the progress of technology and computing, that is now available in the acoustics arena, will provide a strong capability to allow the enhancement of operational crew performance by the use of the auditory mode as a synergistic supplement to the more heavily utilised visual senses.

6. REFERENCES

[1] Transmission and Receptions of sounds under combat conditions Office of Scientific Research and Development, USA. 1946.

[2] German Aviation Medicine, World War II, Dept of the Air Force USA. Vols 1 & 2. 1950.

[3] Harpur, K. "A Noise Survey Performed in the Hercules C1/3." DRA/AS/MMI/CR94189/1. 1994.

[4] Hancock, M. "The contribution of communications noise to aircrew overall noise dose." DRA Report in publication. 1996

[5] Glenn,M.C. "Further assessment of headgear attenuation and noise dose in operational Phantom aircraft". RAE Tech.Memo FS52. Oct 1975.

[6] Hancock, M. "A Report on Aircraft Noise Surveys completed to date". DRA/AS/MMI/CR96123/1. 1996.

[7] James,S.H. "An investigation of the accoustic attentuation of oxygen mask and related aspects. RAE TR91001. Jan 1991.

[8] Rogers, I. "An Assessment of the Benefits Active Noise Reduction Systems provide to speech intelligibility in Aircraft Noise Environments." DRA/AS/MMI/TR96015/1. Feb 96.

[9] South, A. "Final Report of the TV-Tabs/DVI Tornado Trials." DRA/AS/MMI/94066/1. March 94.

[10] Beranek,L.L. (Ed). Noise and vibration control. Academic Press, New York. 1970.

[11] Lower, M. et al. "The design and production of audio warnings for helicopters ISVR". Report AC 527A. October 1986.

[12] James, S.H. "The development of an integrated warning for military helicopters and the initial flight trials in Lynx ZD285." RAE Tech Memo FS1006. 1990.

[13] Mungur,S.H. Current and future requirements for audio threat warnings in military helicopters. DRA/AS/FS/CR93007/1. June 1993.

[14] Munger, S.H. "Audio warning signals in the Harrier GR5." DRA Tech Memo GS1037. August 1992.

[15] Lucas, S.H. and Rood, G.M. "Evaluation of an Active Noise Reduction System for the Mk4 flying helmet during laboratory and helicopter trials". RAE Tech. Report TR85035. April 1985.

[16] Wheeler,P.D. and Halliday, S.G. "An active noise reduction system for aircrew helmets. AGARD CP311 1981.

[17] AGARD CP-311. Aural communication in aviation. 1981.

[18] James,S.H. and Harpur,K.M. In-flight assessment of a helmet mounted active noise reduction system in Sea Harrier FRS1. RAE Tech Memo MM33. January 1990.

[19] James,S.H. and Rood,G.M. In flight assessment of an ANR system in the Sea King 5 helicopter. RAE Tech Memo. MM15. 1989.

[20] James, S.H. "Targets for Flight Helmet Noise Attentuation and Noise Levels at Aircrews Ears." DRA/AS/MMI/CR95033/1. January 1995.

AUDIO DISPLAY TECHNOLOGY

Richard L. McKinley

Bioacoustics and Biocommunications Branch
Armstrong Laboratory, AL/CFBA
2610 Seventh Street
Wright-Patterson Air Force Base, Ohio 45433-7901, USA

1. SUMMARY

The scientific community has experienced substantial growth in knowledge and in the understanding of human auditory localization, particularly in recent decades. This background has spawned the concept of 3-dimensional (3-D) sound and has demonstrated that audio cues can be created and presented over headphones that indicate the location of sounds around the listener. This concept has been incorporated in prototype and commercial systems that synthetically create this virtual or 3-dimensional audio display. Spatial auditory information via 3-D audio displays, has demonstrated significant enhancements in target detection and acquisition, threat avoidance, voice communications enhancement, and situational awareness in laboratory investigations, simulators, and flight demonstrations. Numerous applications in both military and civilian arenas have been identified, and many demonstrated. Although significant enhancements have been obtained, ongoing work is required in the areas of display resolution, head related transfer functions with an emphasis on elevation cues, spatial auditory symbology, distance cues, and sensory interactions involving audio/visual and audio/visual/vestibular systems. Research and development will continue to enhance the understanding and performance of 3-D audio displays. Applications of this spatial auditory information technology will continue to expand in all areas providing even greater increases in user performance and safety.

2. INTRODUCTION

Fifteen years ago, at the last AGARD meeting on aural communications in Soesterberg, Netherlands, three papers were presented on audio and voice warnings. Since that meeting an exciting new auditory display technology has been successfully developed, 3-dimensional (3-D) audio displays. Most experts in 1981 did not believe that auditory localization of signals presented via headphones was possible, yet today, test pilots are flying high performance fighter test aircraft with flight worthy 3-D audio display systems. At this AGARD meeting on audio effectiveness in aviation, six papers on 3-D audio technology are being presented along with three papers on auditory warning signals. The development of this new 3-D audio technology has numerous potential applications in both aviation and ground based environments and has great promise to dramatically enhance the way audio information is presented and used in aviation.

3. BASIC CONCEPT

The basic concept of 3-D audio displays is to create sounds for presentation over headphones which contain information that indicates the location of the sounds around the listener. The location of the sound is

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

perceived to be stationary even while the listener is moving her/his head and looking around. A unique feature is the perceived location of the sounds outside the head of the listener.

Recall that lateralization occurs when a signal in either channel of a binaural headset is presented earlier or louder than the signal in the other channel. The listener perceives the sound as located at the ear receiving the earlier or louder sound or somewhere between the ears, while it remains inside the head. 3-D audio sounds, or virtual audio, refers to synthesized audio signals that are associated with perceptions that place them at specific locations outside the head of the listener.

3.1 Natural Localization

Auditory localization occurs at all times and is innate, natural, easy, unconscious behavior. This localization ability is derived primarily from the differences at the two ears between the time, intensity, and spectral characteristics of the acoustic signals. This cueing information is translated or interpreted by the nervous system as originating at a location in space around the listener.

A simple explanation of how auditory localization in the real world works with the two ears is related to the relative distances of the ears from the sound source. A sound not equidistant from both ears reaches the closer ear in less time and at higher amplitude than at the further ear. The difference between the signal arrival time at the two ears is defined as the interaural time difference (ITD). The head also produces a shadow effect reducing the level of the sound at the more distance ear. In addition, the acoustic signal spectrum changes according to specific locations in space around the head. This spectral change, which can be described as a transfer function, is also modified by the head, torso, and pinnae reflections and it too, is different at the two ears. These spatially correlated changes in sound signal spectrum are called head related transfer functions (HRTFs). Currently it is believed that listeners use a combination of timing and spectral cues along with head or sound source movement to determine the location of the sound source.

3.2 Synthesized Localization (3-D or virtual audio)

The concept for 3-D audio is to synthesize audio localization cues artificially placing a sound at a specific location in the space around a listener. The method is to measure HRTFs and ITDs and use them to create synthetic cues that will provide location information to any input sound of choice. These cues must be correlated with head position/movement for the perceived signal to be accurately localized and externalized. Many 3-D synthesis systems currently use magnetic headtracking systems to measure head position/movement and provide that information to the 3-D audio synthesis system.

3.3 Localization Cue Development

Auditory localization cues for the Armstrong Laboratory Auditory Localization Cue Synthesizer (ALCS) were created from ITD and HRTF measurements in the Auditory Localization Facility (ALF). ALF is a 14 ft diameter geodesic sphere that completely surrounds the subject a full 360 degrees. The sphere, which is housed in a large anechoic chamber, contains 277 loudspeakers, one located at every node (hub) or intersection point of the sphere. The acoustic manikin or human subject is positioned inside the sphere with the head positioned in the center of the sphere. Miniature microphones are placed in each ear of the manikin or subject and the

sphere. Miniature microphones are placed in each ear of the manikin or subject and the ITDs and HRTFs are measured. These HRTFs and ITDs are then modeled using digital filters which are convolved with the input signal using a digital signal processor. The resulting signal is presented via headphones to the listener.

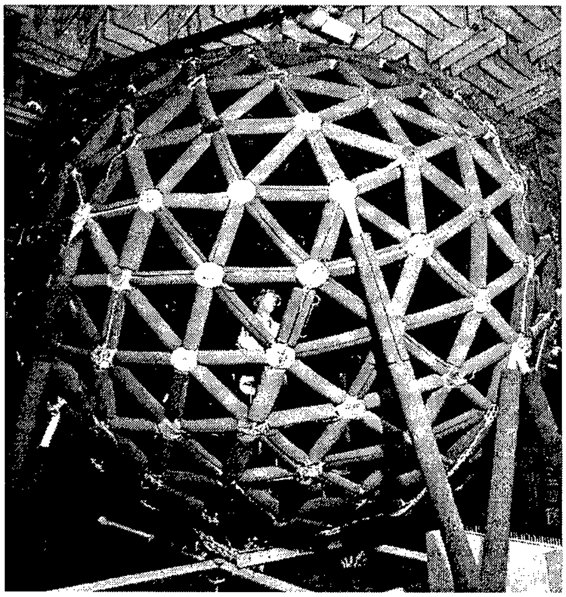


Figure 1. Auditory Localization Facility (ALF) inside Large Anechoic Chamber with an Acoustic Manikin at the Center Position inside the Sphere

3.4 3-D Audio /Stereo/Surround Sound

The production and perception of 3-D audio differ from those of stereophonic and surround sound technologies. 3-D audio displays have been successfully demonstrated using headphone presentation. The concept of 3-D audio is to generate a signal that is perceived to be emanating from a specific location. The concept of surround sound is to generate a signal which seems to be coming from everywhere, "surrounding" the listener. Stereo imaging seems to come from the left or right loudspeaker or somewhere between them. Surround sound is perceived by many to be more desirable than stereophonic sound. Neither stereophonic sound nor surround sound is capable of providing audio cues with the unique spatial localization capabilities experienced with 3-D audio displays over headphones.

4. POTENTIAL APPLICATIONS

4.1 Military Applications

A major application of 3-D audio display technology in the military is in improving situational awareness. Aviator attention must be distributed across many activities in complex cockpits with high workloads at critical times in such situations as combat maneuvering, bad weather, night flying, navigation, mission tactics, and many more. Applications of 3-D audio in several of these areas, and others, should significantly increase situational awareness and enhance both safety and performance.

Some specific applications include 1) threat cueing with the Radar Warning Receiver, 2) target location cueing, 3) navigation 4) collision avoidance cueing, and 5) wingman location cueing. Voice communication enhancement is provided when multiple radios are virtually separated in space around the listener using the 3-D audio. Visual performance with night-vision systems where the field of view is limited is also a strong candidate for integration with 3-D or virtual audio displays. The list of potential applications is quite long and it continues to grow.

4.2 Civilian Applications

There are many applications of 3-D audio that are appropriate for transition to the civilian sector. Among them is the use of spatial

audio display information to enhance the situational awareness of pilots in instrument meteorological conditions is very promising, as well. The demonstrated advantages of virtually separated communications merged with the location in space information should significantly increase the performance of dispatching for police, taxi, and fire fighters as well as for ground control in aircraft terminals. Perhaps the most widespread applications of 3-D audio are and will be in the areas of education, training, and entertainment. The possible refinements in education and training processes provided by this technology, including personal virtual experiences associated with the instruction, are still in very early stages of development. Possibly, in the near future, all interactive entertainments will include 3-D audio, and those without this technology will become obsolete.

5. COMMERCIAL SYSTEMS

Virtual or 3-D audio display systems are commercially available. Some of the manufacturers provide several models, typically representing different degrees of resolution and other options. Perhaps the most widely recognized systems are those of Scott Foster, Crystal River Engineering, and his first model called the Convolvotron. Two of the more recently offered 3-D or virtual audio systems are the DAC-1 from Tucker Davis Technologies and the 3-D Geni from Systems Research Laboratories.

State of the art 3-D audio systems include a sampling rate of 44.1 KHz, 16 bits, and real time computing with 120 to 1024 TAP FIR HRTF Filters. Spatial HRTF samples range from 1 to 18 degrees with real time head position updates.

6. BACKGROUND

Now that you have an idea of the concept, technology, and potential applications, let me describe some of the background of 3-D audio technology.

The scientific community is attempting to converge on a unifying model of human auditory localization. The evolution of a unifying model has been taking place for over 100 years. Fechner, in 1860, was one of the earliest researchers of the phenomenon of human auditory localization. Batteau reported in 1963, on a time delay theory of auditory localization. Blaurt, in 1969-70, found that sounds falling on the pinnae, head, and ear canal were modified according to the angle of arrival of the sound to the ear and that these changes were frequency dependent. This resulted in a model of localization based on timbre references. Shaw has probably done the most extensive work on understanding the effects of pinna structure on auditory localization. The combination of these theories resulted in what is currently described as the duplex theory of auditory localization. In the duplex theory, it is proposed that the listener uses both interaural time differences and interaural intensity differences to determine sound source location.

Burkhard, in 1975, described the development and characteristics of an acoustic manikin created in an attempt to accurately simulate the acoustic diffraction of the head and torso and include the effect of the pinna and ear canal. This Knowles Electronic Manikin for Acoustic Research (KEMAR), has been extensively used by researchers investigating auditory localization. However, listening to binaural auditory signals from an acoustic manikin (i.e., a binaural recording) does not provide the dynamic acoustic cues that allow the listener to localize the sound source in exocentric space. In 1974, Lambert proposed a dynamic theory of auditory localization

based on the effects of head movement. He created a localization measurement facility configured around a KEMAR manikin surrounded by a series of loudspeakers arrayed in the horizontal plane. With this apparatus and adjustments to the position of sound relative to the manikin to match the head motions of the listener, Doll was able to demonstrate auditory localization over headphones. Doll reported that for localization in azimuth, interaural time delays and head motion were two critical parameters.

In 1988, McKinley described a concept for a localization cue synthesizer that included the three parameters of interaural time delays, head movement, and pinnae, head, torso, and ear canal transforms [head related transfer functions or HTRFs as described by Blauert in 1983]. About the same time, Beth Wenzel of NASA Ames Research Center began development of a system with the same objectives as the Armstrong Laboratory system. This development effort resulted in the system cited earlier called the Convolvotron, produced by Scott Foster. Both the NASA Ames and the Armstrong Laboratory systems are currently pursuing improvements in performance and applications.

Much research by Wightman and Kistler has focused on interaural time delay information and HRTFs. These researchers, working at the University of Wisconsin under sponsorship of NASA-Ames and the Armstrong Laboratory, found that the interaural time information is most critical for localization in azimuth. At the Armstrong Laboratory, confirmation was provided that generic synthesized localization cues created from HRTFs measured on an acoustic manikin were effective localization cues for the average observer. However, the HRTFs measured on individuals differ significantly from one person to another. Consequently, systems that generate synthesized 3-D audio displays must model the HRTFs of an individual for that person to obtain optimum performance, as reported by Wenzel in 1993.

During 1991, the Armstrong Laboratory designed and developed a flight-worthy version of the auditory localization cue synthesizer (ALCS) 3-D audio display generator or 3-D Gen. Later in 1991, DARPA funded Armstrong Laboratory to develop an integrated audio helmet that included 3-D audio displays, active noise reduction headsets, an advanced noise canceling microphone, head tracking sensors, and physiological monitoring. This integrated audio helmet was to be used for demonstrations and performance data collection in high fidelity flight simulators. This DARPA sponsored effort produced a very successful lightweight integrated flight helmet. In 1992, DARPA sponsored a flight demonstration of this system on a U.S. Marine Corps AV-8B "Harrier," a two-place cockpit aircraft that had been previously modified to include a militarized head tracking system. This was the first dlight demonstration of 3-D audio displays.

6.1 Summary of Flight Demonstration

A target acquisition and verification task was selected for the initial flight demonstration. Navigation points or targets were positioned on the ground in various groupings at different locations along the flight route. Typically, information from these targets or navigation points appears as visual displays on flat screens in the cockpit. The aviator must first look at the visual display and then through the windscreen to acquire and verify the target. The aircraft was flown through scenarios that required the crew to identify a single target

from each of the groups of targets on the ground. The 3-D audio display cues from the system installed in the aircraft were provided only to the aviator in the rear seat. The aviator in the front seat selected one of the targets in each group for identification. The aviator in the rear seat did not see any of the visual displays or targets and groupings through the windscreen. 3-D audio cues from the synthesizer were added to the selected target warning signal and presented over headphones to the aviator in the rear seat. That aviator's task was to identify the selected target in the group using only the 3-D audio signals.

The aircraft approached the targets at high speeds. The 3-D audio display system worked well during the demonstration. Aviators had no problem identifying targets separated by 20 degrees. Most could discern targets separated by as little as 12 degrees. The system provided azimuth cues reliably to approximately half a clock code (15 degrees). 3-D elevation cues did not work as well as azimuth cues. Aviators expressed confidence in target location judgements as being either low or high in elevation. Elevation cues improved during steep angles of bank. Voice communications separation by the virtual location of the cockpit radios to 45 degrees right and left of the aviator worked very well. One aviator commented that only by using the separation feature was he able to accurately copy the dual message traffic. Overall, the flight demonstration verified the high value added merits of the 3-D audio, identified some limitations, and provided direction for ongoing and future initiatives with this new technology.

Auditory displays now include not only the traditional auditory and/or voice warnings but also this new method of presenting information in the form of 3-D or virtual audio displays. There are significant advances in the state-of-the-art in both general areas. The rapidly expanding area of 3-D audio displays is lacking in basic auditory symbology as well as that associated with particular display applications. During this AGARD conference, only one paper was presented that addressed spatial or 3-D audio symbology. This is one area that probably deserves much more research focus than it has been given in the past.

7. RESEARCHERS

Numerous researchers in many countries are investigating auditory localization and 3-D audio displays. The recalled researchers listed below are those with whom I have had contact and am familiar with their work. The reason for including this list is to demonstrate the broad interest and focus in this new technology area and not to identify individuals or countries. Several of these researchers are presenting papers at this conference in topic areas ranging from basic science in auditory localization to applications of 3-D audio displays in flight test aircraft. Please forgive any omissions.

Dennis Folds	USA
Russell Martin	DSTO, Australia
Jens Blauert	Germany
Roy Patterson	Cambridge, UK
Lionel Pellieux	CERMA, France
Albert Bronkhorst	Netherlands
Henrik Moller	Denmark
Soren Bech	Denmark
Simon Oldfield, Simon Parker	Australia
Elizabeth Wenzel, Derand Begault	NASA
Kim Abouchachra	Army, USA
Tom Buell	Navy, USA
Fred Wightman, Doris Kistler	University of Wisconsin
Robert Gilkey	Wright State University

Richard McKinley, Mark Ericson, William D'Angelo & Bart Elias USAF

8. CURRENT PERFORMANCE

Information on human performance with the 3-D audio display system has been obtained from various laboratory studies and from measurements of individual auditory localization ability on over 200 subjects. The current state-of-the-art performance in azimuth with generic HRTFs and when subjects are permitted to move their heads is approximately 4 to 5 degrees. Elevation resolution with generic HRTFs is approximately 25 to 35 degrees and it improves with custom HRTFs to 10 to 15 degrees. Speech intelligibility improvements of 25 to 35 percent are obtained by virtually separating multiple talkers with an equivalent improvement of 3 dB speech signal-to-noise ratio for a single talker. Audio aided search experiments have shown an average 50 percent decrease in target acquisition times and a 50 to 100 percent improvement in target detection range with the addition of spatial auditory cues.

9. FUTURE DIRECTIONS

The amount and variety of work that needs to be accomplished is sufficient to keep all interested researchers busy for the foreseeable future. Some future directions have emerged from investigations and experiences with 3-D audio displays in laboratories and field operations. The area requiring the most attention is auditory symbology. The development of spatial auditory symbology (and auditory icons) must continue with an emphasis on environments in which multiple 3-D audio symbols will be presented simultaneously. Although various applications may require different levels of resolution, spatial resolution of synthetic 3-D audio displays is 3 to 5 degrees while free-field resolution is 1 to 2 degrees. Ongoing work is required in the definition, acquisition, and application of HRTFs, particularly as they influence elevation performance.

Another area showing increasing importance and impact is that of sensory interaction. Initial efforts in audio/visual interactions should be continued and followed with audio/visual/vestibular interactions in the 3-D audio localization function. Distance cues will not be required in many 3-D audio applications, however they are very important to those applications in which distance information is critical.

Research and development should continue in laboratories, simulators, and in-flight studies to enhance the understanding and performance base of 3-dimensional auditory displays and their interaction with other related technologies such as helmet mounted displays. Applications of this technology will continue to expand in both military and civilian environments to improve situational awareness, user performance, and to increase safety.

The Effects of Spatial Auditory Preview on Visual Performance

Bartholomew Elias
US Air Force Armstrong Laboratory
Noise Effects Branch
AL/OEBN
2610 Seventh Street
Wright-Patterson AFB, Ohio, USA 45433-7901

1. SUMMARY

Since the auditory system is not spatially restricted like the visual system, spatial auditory cues can provide information regarding an object's position, velocity, and trajectory beyond the field of view. Recent studies (e.g., Perrott, Cisneros, McKinley, & D'Angelo, 1995) have demonstrated performance benefits in static visual search tasks over large spatial extents when visual targets have been augmented with spatial auditory position cues. The benefits of spatial auditory display augmentation have also been demonstrated in applied settings such as airborne traffic collision avoidance systems (Begault, 1993). Research has also shown that spatial auditory displays are potentially useful for enhancing cockpit situational awareness and reducing visual workload in tactical aircraft operations (McKinley, et al., 1994). The research program described here adds to these initial findings regarding the utility of spatial auditory displays by demonstrating that visual displays can be augmented with dynamic spatial auditory preview cues that provide information regarding the relative position, velocity, and trajectory of objects beyond the field of view. In one experiment, the effects of a spatial auditory preview display were examined in a visual target aiming task. A moving sound source provided cues regarding the position and velocity of moving targets prior to their appearance on the visual display. By providing these spatial auditory preview cues, greater accuracy was achieved in the visual target aiming task. In a second experiment, dynamic spatial auditory cues presented through headphones conveyed preview information regarding target position, velocity, and trajectory beyond the field of view in a dynamic visual search task. The provision of spatial auditory preview cues significantly reduced response times to acquire and identify moving visual targets that traversed a cluttered display and significantly reduced error rates in target classification. These findings demonstrate that spatial auditory preview can augment visual displays and enhance performance in complex, dynamic task domains such as aviation.

2. INTRODUCTION

Recent advances in auditory display technology have made possible the real-time presentation of spatial sounds through headphones by using digital filtering techniques to replicate key auditory spatial cues (see, e.g., Wenzel, 1991). These auditory spatial cues consist of interaural time differences (ITDs), interaural intensity differences (IIDs) and distortions of the acoustic signal created by the pinna or outer ear and the upper torso. Spatial auditory cues may also include reverberations, echoes and other signal distortions generated in specific acoustic environments. In dynamic environments, time-varying characteristics of the acoustic signal such as frequency dependent changes in sound level and *Doppler* shifts can be incorporated to provide cues regarding the relative motion of a sound source. By digitally creating these cues, an auditory signal presented through headphones can convey spatial information regarding the position and movement of a *virtual* sound source.

In response to these technological advances, a line of research has emerged exploring fundamental issues regarding the possible domains of application and potential utility of this technology. Since the auditory system is not spatially restricted like the visual system, the auditory modality can provide spatial information over the full 360 degrees in both azimuth and elevation. In this manner, spatial auditory presentations can provide cues regarding the position and movement of unseen objects. Furthermore, spatial auditory stimuli can provide critical directional cues to orient us toward objects in the periphery and outside the field of view, thereby enabling rapid visual acquisition of these objects and timely execution of visually guided motor responses.

Studies exploring the use of spatial auditory displays for augmenting static visual searches over large spatial extents have demonstrated that providing spatial auditory position cues can significantly enhance one's ability to rapidly detect and identify visual targets (see Perrott, Saberi, Brown, & Strybel, 1990; Perrott, Sadralodabai, Saberi, & Strybel, 1991; Perrott, et al., 1995, Strybel, Boucher, Fujawa, & Volp, 1995). The benefits of spatial auditory display augmentation have also been demonstrated in applied settings such as airborne traffic collision avoidance systems (Begault, 1993; Begault, 1995), low visibility aircraft ground operations (Begault, 1995), and as a potential means for enhancing cockpit situational awareness and reducing visual workload in tactical aircraft operations (McKinley, Ericson, & D'Angelo, 1994).

The aforementioned research focused specifically on the role of static spatial auditory cues in conveying information regarding the position of visual targets. However, in highly dynamic task environments such as aviation, dynamic spatial auditory cues can be utilized to convey information regarding the velocity, trajectory, and instantaneous position of moving objects in the periphery and beyond the field of view. A set of laboratory experiments was conducted to evaluate the performance

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

benefits of providing dynamic spatial auditory preview cues regarding the position, velocity and trajectory of targets beyond the field of view in two visually guided tasks: 1) dynamic visual target aiming and 2) dynamic visual search.

3. EXPERIMENT 1, VISUAL TARGET AIMING

3.1 Subjects

Eight right-handed males aged 19 to 27 served as subjects. All subjects had normal hearing and corrected far-field visual acuity of 20/20 or better. Subjects received ten dollars for each of ten experimental sessions. Furthermore, the best performer in each of two groups of four subjects received an additional twenty dollars as a performance incentive.

3.2 Apparatus

Experimental sessions were conducted in an acoustically controlled experimental chamber (see Figure 1). The subject was seated centrally within the chamber facing a large black fabric screen located at a distance of 150 cm. A 3000 mm linear slide powered by a stepper motor was located behind the screen and out of the subject's view. A dynamic sound source was created by mounting a small speaker to the carriage of the linear slide. The auditory stimulus presented though this dynamic sound source was a dual one-third octave band-filtered noise centered at 400 Hz and 2,500 Hz. Signal presentations were synchronized with the motion of the linear slide and ranged in intensity from 76 dB(A) to 81 dB(A) at the subject's location as a function of sound source distance. Ambient pink noise was presented continuously at a level of 76 dB(A) through a speaker located directly behind the subject. The computer generated visual display consisted of two small white squares representing a target and a projectile displayed on a black background (see Figure 2). This visual image was projected onto the black fabric screen using an LCD projection panel and measured 160 cm by 120 cm. Subjects oriented their gaze toward the center of the display and a chin rest was used to prevent head movements.

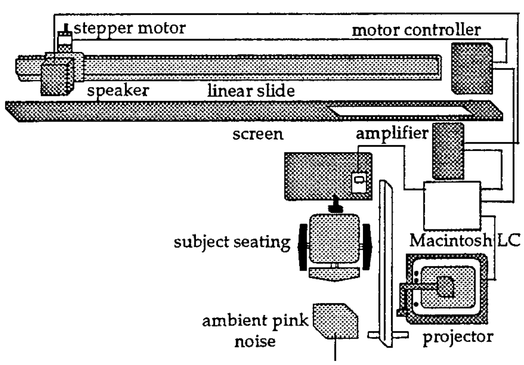


Figure 1. Layout of the experimental test chamber.

3.3 Experimental Task

Subjects performed the target aiming task by pressing the mouse button to fire the projectile at the visual target. Depressing the mouse button initiated movement of the projectile which moved at a constant velocity of 40 cm/s. The task required precise timing of responses to achieve the coincident arrival of the target and the projectile at their point of intersection on the visual display. The velocity of the target on each trial was randomly chosen from a set of three constant speeds (44 cm/s, 64 cm/s, and 84 cm/s) as a manipulation of task difficulty. Relative error magnitude (measured as the distance between the target and the projectile when the projectile crossed the target's path) served as the dependent variable. Through systematic manipulations of the position and velocity of the sound source and the visual target, the effects of dynamic auditory preview were assessed.

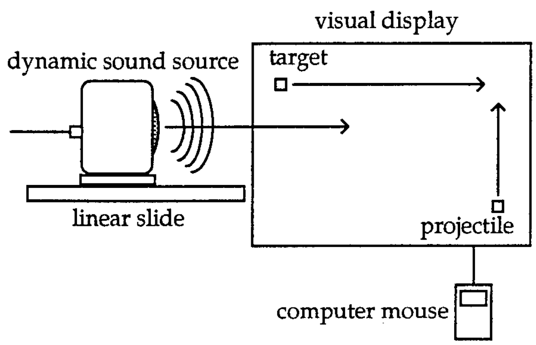


Figure 2. The target aiming task.

3.4 Procedures

3.4.1. Phase I, Visual Training

On the first day of the experiment, each subject completed 300 practice trials using only visual spatial cues at three different target speeds: 44 cm/s, 64 cm/s, and 84 cm/s. During this training phase, the auditory stimulus remained stationary behind the visual display and was onset two seconds prior to the presentation of the visual target.

3.4.2 Phase II, Aligned Auditory Preview

On Days 2 through 6 of the experiment, subjects completed 300 trials per day using auditory preview cues provided by motion of the sound source beyond the bounds of the visual display that was in exact alignment with the position and velocity of the upcoming visual target. This experimental phase consisted of a 3 (target velocity: 44 cm/s, 64 cm/s, or 84 cm/s) by 4 (auditory preview distance: none, 100 cm, 180 cm, or 260 cm) repeated measures design.

3.4.3 Phase III, Misaligned Auditory Preview

On Days 7 through 10 of the experiment, the effects of misalignments between the position and velocity of the sound source and the visual target were assessed. In this phase, subjects were assigned to one of two experimental groups. Subjects in Group A completed trials having position mismatches between the auditory source and the visual target while subjects in Group B completed trials having velocity mismatches between the auditory source and the visual target. The first 60 trials presented each day during this phase were identical to trials presented during Phase II in order to maintain the informational relevance of dynamic sound source.

3.4.3.1 Group A, Position Mismatches

The four subjects assigned to Group A completed the remaining 240 trials each day at two different auditory preview distances (100 cm, 260 cm) with two target velocities (44 cm/s, 84 cm/s) under three different position mismatch conditions. In position mismatch conditions, the auditory source was displaced either 75 cm to the right of the visual target (+75 cm) or 75 cm to the left of the visual target (-75 cm). In the control condition, there was no misalignment between the position of the auditory source and the visual target. Thus, the experimental design of Phase III for Group A subjects consisted of a 3 (position mismatch: -75 cm, 0 cm, +75 cm) by 2 (auditory preview: 100 cm, 260 cm) by 2 (target velocity: 44 cm/s, 84 cm/s) repeated measures design.

3.4.3.2 Group B, Velocity Mismatches

The four subjects assigned to Group B completed the remaining 240 trials each day at two different auditory preview distances (100 cm, 260 cm) with two target velocities (44 cm/s, 84 cm/s) under three different velocity mismatch conditions. In velocity mismatch conditions, the auditory source moved at a rate either 20 cm/s slower than the visual target (-20 cm/s) or 20 cm/s faster than the visual target (+20 cm/s). In the control condition, the auditory source and the visual target moved at the same rate. Thus, the experimental design of Phase III for Group B subjects consisted of a 3 (velocity mismatch: -20 cm/s, 0 cm/s, +20 cm/s) by 2 (auditory preview: 100 cm, 260 cm) by 2 (target velocity: 44 cm/s, 84 cm/s) repeated measures design.

3.5 Results

3.5.1. Phase II, Aligned Auditory Preview

Mean relative error magnitudes for Phase II trials are shown in Figure 3. A repeated measures ANOVA revealed a significant auditory preview distance by target speed interaction, $\underline{F}(6,42) = 44.62$, $\underline{p} < .001$. In the critical test condition where the targets moved at 84 cm/s and there was insufficient time to make accurate firing responses using only visual cues, performance improved significantly with increasing auditory preview distance, $\underline{F}(3,21) = 63.74$, $\underline{p} < .001$. In conditions where the target moved at 44 cm/s or 64 cm/s, sufficient time was available to respond using the available visual cues and the addition of spatial auditory information had a negligible effect on performance. These results indicate that the provision of dynamic auditory preview can aid task performance when visual cues alone are insufficient for making accurate responses.

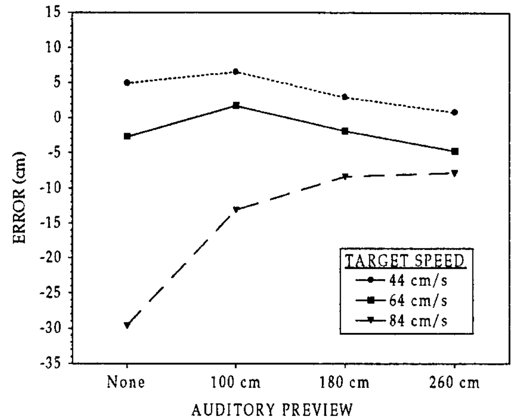


Figure 3. Mean relative error magnitudes as a function of auditory preview distance and target speed.

3.5.2 Phase III, Misaligned Auditory Preview

3.5.2.1. Group A: Position Mismatches

Mean relative error magnitudes for position mismatch trials are shown in Figure 4. A repeated measures ANOVA indicated a significant two-way interaction between position mismatch condition and target speed, $\underline{F}(2,6) = 137.97$, $\underline{p} < .001$. No effect for position mismatch was found among trials where the target moved at 44 cm/s. However, a significant effect for position mismatch was demonstrated among trials where the target moved at 84 cm/s and insufficient time was available to make accurate responses using only visual cues, $\underline{F}(2,6) = 150.37$, $\underline{p} < .001$.

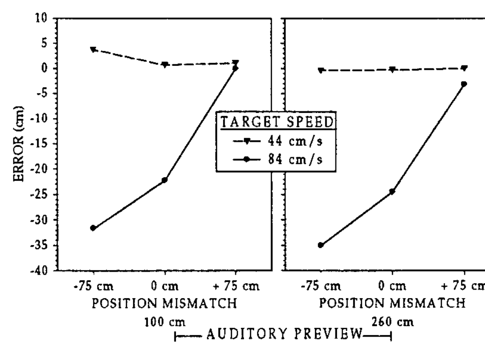


Figure 4. Mean relative error magnitude as a function of position mismatch, target speed and auditory preview distance.

Tukey post hoc comparisons indicated that trials where the sound source was displaced 75 cm to the left of the target (-75 cm) produced significantly larger error magnitudes than the control condition. However, trials where the auditory source was displaced 75 cm to the right of target (+75) actually resulted in significantly lower error magnitudes than the control condition. These results demonstrate that a dynamic auditory preview display that lags behind its visual correlate may disrupt performance. However, results suggest that a dynamic auditory preview display that precedes its visual correlate may actually enhance performance presumably by compensating for inherent perceptual-motor delays in visual responding.

3.5.2.2 Group B: Velocity Mismatches.

Mean relative error magnitudes for velocity mismatch trials are shown in Figure 5. A repeated measures ANOVA revealed a significant main effect for velocity mismatch condition, $F(2,6) = 29.55$, $p < .001$. These results suggest that auditory preview cues can prime visual responses. Specifically, when a relatively faster sound source preceded the fast target or when a relatively slower sound source preceded the slow target, uncertainty was reduced and performance improved compared to the control condition. However, when a relatively slower sound source preceded the fast target or when a relatively faster sound source preceded the slow target, subjects were misled by the auditory preview and performance consequently suffered.

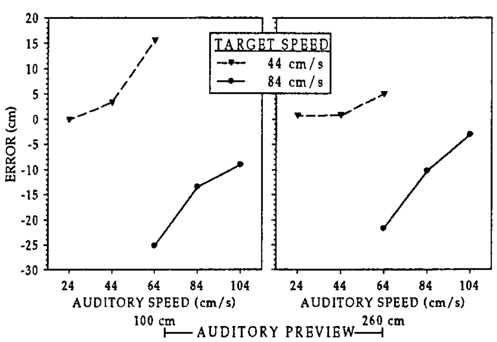


Figure 5. Mean relative error magnitude as a function of target speed, sound source speed, and auditory preview distance.

4. EXPERIMENT 2, DYNAMIC VISUAL SEARCH

4.1 Subjects

Eight right-handed subjects (5 male and 3 female) ranging in age from 19 to 30 participated in this experiment. The subjects had normal hearing and corrected far-field visual acuity of 20/20 or better. Subjects were recruited from a panel of research subjects retained by an on-site contractor for participation in various psychoacoustic studies. Subjects were paid for their participation at a rate commensurate with their length of service on the subject panel. In addition, the best performer on the experimental task was awarded an additional twenty dollars as a performance incentive.

4.2 Experimental Environment

Experimental sessions were conducted in an acoustically controlled experimental chamber measuring 13 feet by 15 feet. The subject was seated centrally within the room facing a large display panel from a distance of 150 cm (see Figure 6). Computer generated graphics were projected on to the display panel using an LCD projection panel illuminated by an overhead projector. The projected image subtended a visual angle of 56 degrees in azimuth and 50 degrees in elevation at the subjects location.

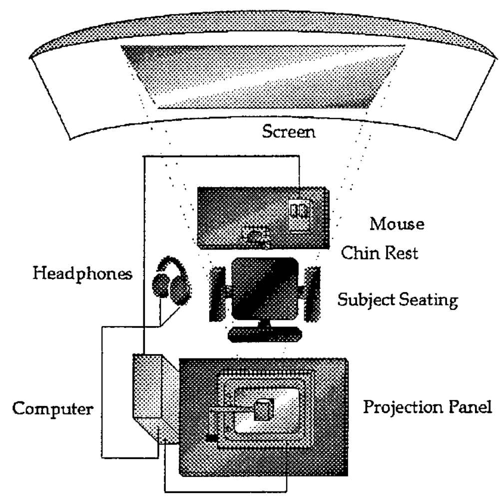


Figure 6. Layout of the experimental chamber.

4.3 Experimental Task

In order to assess the effects of auditory preview on dynamic visual target identification, a two-alternative forced choice (2AFC) dynamic visual search task was implemented in this experiment. Subjects were required to search among a group of moving distractors to acquire the target symbol and identify it as either a "FRIEND" or an "ENEMY" in accordance with the categorization scheme shown in Figure 7. Each symbol subtended approximately 5.73 degrees of visual angle in both

azimuth and elevation. The symbols consisted of 50% gray-scale pixels and were presented on a black display background.

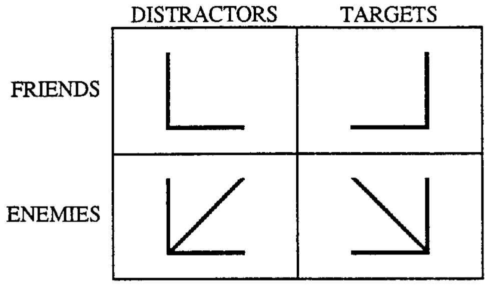


Figure 7. Symbol categories used in the experimental task.

On a given trial, either 12, 24, or 32 distractors moved to and fro across the display screen as a manipulation of visual display load. Distractors initially appeared at random locations on the display screen and moved along one of eight linear trajectories reversing their direction when they reached the edge of the visual display. On a given trial, all of the distractors traveled at the same rate. However, on half of the trials, the rate of motion was 3 degrees per second (deg./s), and on half of the trials, the rate of motion was 6 deg./s. In each of these instances, half of the distractors were "ENEMIES" and half of the distractors were "FRIENDS".

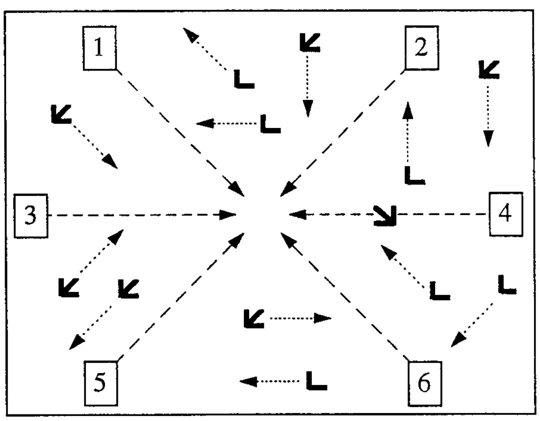


Figure 8. Possible motion trajectories of the target and distractors. The above scene depicts 12 distractors traveling to and fro and an "ENEMY" target traversing the screen on trajectory # 4. (Note: trajectory lines and numbers were not visible on the display screen).

After a random delay period ranging from 10 to 20 seconds, the target appeared. On half of the trials the target was a "FRIEND" and on half of the trials the target was an "ENEMY". The target originated at one of six randomly determined points along the border of the visual display and traveled along the corresponding linear trajectory as indicated in Figure 8. The target traveled at the same rate as the distractors. Thus, the target traveled at 3 deg./s on half of the trials and at 6 deg./s on half of the trials. On half of the trials, spatial auditory cues were presented to subjects prior to the appearance of the visual target. This spatial auditory preview was synchronized with the movement of the yet to be seen visual target and conveyed information regarding target position, velocity and trajectory. On the remaining trials, the auditory stimulus was presented monaurally and was onset 2.5 seconds prior to the appearance of the visual target. Thus, the monaural auditory cue served as a warning signaling the target's approach, but conveyed no position, velocity, or trajectory information. By comparing trials in which spatial auditory cues were presented to trials in which only the monaural stimulus was presented, the effects of spatial auditory preview on visual search performance was assessed.

4.4 Spatial Auditory Stimuli

The spatial auditory stimuli consisted of a set of twelve virtual moving sounds. These stimuli consisted of digital recordings of a click train (specifically, a 7 Hz square-wave tone) that moved along one of six trajectories corresponding to the six possible motion trajectories along which the target could travel. These trajectories extended to 60 degrees left and right of center, thus providing dynamic spatial cues that spanned 32 degrees beyond the bounds of the visual display. These dynamic spatial auditory stimuli were recorded at two different apparent motion velocities: 3 deg./s and 6 deg./s.

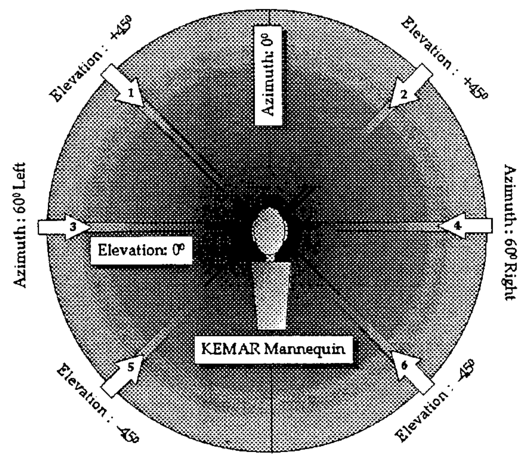


Figure 9. Apparent motion trajectories of the dynamic spatial auditory stimuli.

These acoustic stimuli were produced using a unique facility developed for spatial auditory research. This facility, the Auditory Localization Facility (ALF), located in an anechoic chamber at the Biocommunications Laboratory at Wright-

Patterson AFB, consists of an array of 272 speakers situated in a geodesic sphere arrangement measuring 14 ft. in diameter. To obtain the recorded virtual sounds for this experiment, a Knowles Electronic Mannequin for Acoustic Research (KEMAR) wearing "average ear" pinna molds was placed centrally within this geodesic sphere. Recording microphones were placed inside the ear canals of the KEMAR mannequin and were input to independent channels of a digital audio tape (DAT) recorder. The sequencing of speaker onsets, offsets, and durations within the sphere was programmed to create apparent motion of the click train along each of the six target paths shown in Figure 9. The digital recordings of these presentations were transferred to a computer hard disk where they were edited and stored as stereo sound files. During experimental sessions, these spatial auditory stimuli were presented to subjects using an Antex SX-12a digital audio adapter board whose signal was amplified and played through stereo headphones.

4.5 Experimental Design

Subjects completed the experiment in individual sessions lasting approximately one hour. Each subject completed one session per day over four consecutive days. On each day, subjects completed 144 dynamic visual search trials that were subdivided into three blocks of 48 trials. Display load was varied by presentation of either 12, 24, or 32 distractors. Trial speed was manipulated by moving the target and distractors at a rate of either 3 deg./s or 6 deg./s. Finally, half of the trials included spatial auditory preview presentations while the remaining half of the trials contained only a monaural warning of the target's approach. Thus, the experiment consisted of a 2 (auditory display condition: spatial, monaural) by 2 (trial speed: 3 deg./s, 6 deg./s) by 3 (visual display load: 12, 24, or 32 distractors) repeated measures design. The dependent measures of task performance included response time, measured as the time elapsed between the appearance of the visual target and the execution of the subject's response, and error rates across experimental conditions.

4.6 Results

Mean response times for each of the experimental conditions are shown in Figure 10. Analysis of variance calculations revealed a significant two-way interaction between visual display load and auditory display condition, $F(2,14) = 7.37$, $p < .01$. Simple main effects indicated that the provision of spatial auditory preview significantly reduced response times across all visual display load conditions. However, the benefit derived from spatial auditory preview increased with increasing visual display load indicating that the provision of spatial auditory preview proved most beneficial when visual task demands were high (see Table 1).

Individual error rates for all monaural and spatial auditory visual search trials are presented in Table 2. Across all experimental sessions, a total of 122 errors were recorded on trials in which spatial auditory preview was provided and 160 errors were recorded on trials in which only the monaural cue was presented. A Wilcoxon matched-pairs analysis of subjects' error rates indicated that significantly fewer errors were made when spatial auditory preview was provided, $z = 2.24$, $p < .05$, again indicating a significant performance benefit derived from spatial auditory preview presentations.

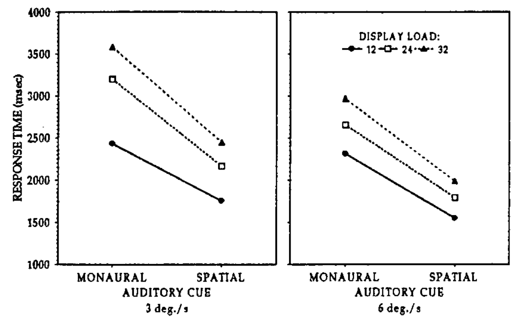


Figure 10. Response time as a function of display load, trial speed and auditory cue condition.

Table 1. Mean Response Times (msec) and Mean Response Time Differences (msec) Between Auditory Display Conditions as a Function of Visual Display Load.

DISPLAY LOAD	AUDITORY DISPLAY MONAURAL	SPATIAL	DIFF.
12	2384.58	1659.11	725.47
24	2929.74	1979.40	950.34
32	3283.38	2216.08	1067.30

Table 2. Individual and total error rates for monaural and spatial auditory visual search trials.

SUBJECT	MONAURAL	SPATIAL	DIFF.
1	20	17	-3
2	4	2	-2
3	18	12	-6
4	18	11	-7
5	14	5	-9
6	8	6	-2
7	57	59	+2
8	21	10	-11
TOTAL	160	122	-38

Finally, an analysis of improvements in response time across trial blocks revealed typical learning curves which are presented in Figure 11. Not surprisingly, response times were significantly reduced through practice, $F(11,77) = 29.34$, $p < .01$. However, no interaction between trial block and auditory display condition was evident, $F(11,77) = 0.35$, $p > .25$, indicating that the beneficial effects of the spatial auditory preview cues were immediately realized and were not dependent on extensive familiarization or training.

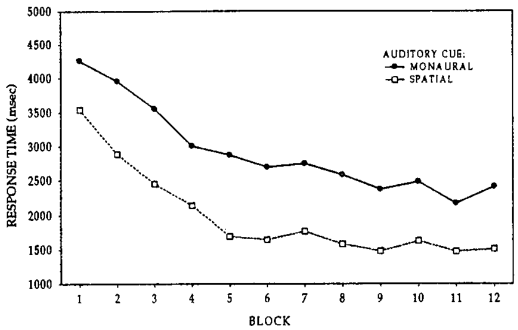


Figure 11. Response time as a function of trial block and auditory cue condition.

5. DISCUSSION

The results of these experiments clearly demonstrate that spatial auditory preview cues can enhance visual performance in target acquisition and aiming. Results from both the dynamic target aiming task and the dynamic visual search task demonstrated that the benefits of providing spatial auditory preview cues are most pronounced when visual task demands are high. In the dynamic target aiming task, the performance benefits derived from spatial auditory preview information were most evident when the target velocity was high and little time was available to respond to the visual stimuli alone. Similarly, in the dynamic visual search task, the performance benefits derived from spatial auditory preview cues were most pronounced when the visual display load was high. These results suggest that spatial auditory displays could provide important benefits in aviation where time critical actions must be executed in response to highly dynamic stimuli presented under conditions of high visual workload.

The findings of this research have also provided insights regarding how spatial auditory displays can be engineered to enhance human performance. For example, in the dynamic target aiming task, subjects achieved greater accuracy when the spatial auditory cue preceded the visual target and thereby compensated for inherent perceptual-motor delays in visual responding. This is analogous to the practice of *quickening* a display, or presenting data regarding the predicted spatial position of display objects at some future time. By designing spatial auditory preview displays in a manner that compensates for inherent perceptual-motor delays in responding to visual stimuli, combined audio-visual spatial display systems can be engineered for enhanced human performance.

The results of these studies have important implications for the implementation of spatial auditory displays in highly dynamic task environments such as aircraft flight decks, air traffic control consoles, and tactical information displays. In these task environments , humans are faced with highly dynamic visual stimuli and extensive demands on visual processing. The findings of this research suggest that virtual auditory preview displays conveying information regarding the movement of peripheral objects can aid visual target acquisition and identification and improve performance in executing visually guided responses. Furthermore, these results indicate that spatial auditory preview displays are a viable mechanism for augmenting visual information in complex dynamic systems. Clearly, there are many potential uses for spatial auditory displays in aviation, aerospace, and military systems. Through detailed research and careful human engineering, the development and implementation of these spatial auditory displays can enhance human performance in these domains of application.

6. REFERENCES

Begault, D. R. (1993). Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation. *Human Factors, 35*(4), 707-717.

Begault, D. R. (1995). Head-up auditory display research at NASA Ames Research Center. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, pp. 114-118.

McKinley, R. L., Ericson, M. A., & D'Angelo, W. R. (1994). 3-dimensional auditory displays: Development, applications, and performance. *Aviation, Space, and Environmental Medicine, May 1994*, A31-A38.

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Perception and Psychophysics, 48*(3), 214-226.

Perrott, D. R., Sadralodabai, T., Saberi, K., & Strybel, T. Z. (1991). Aurally aided visual search in the central visual field: Effects of visual load and visual enhancement of the target. *Human Factors, 33*(4), 389-400.

Perrott, D. R., Cisneros, J., McKinley, R. L., & D'Angelo, W. R. (1995). Aurally aided detection and identification of visual targets. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, pp. 104-108.

Strybel, T. Z., Boucher, J. M., Fujawa, G. E., & Volp, C. S. (1995). Auditory spatial cueing in visual search tasks: Effects of amplitude, contrast and duration. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, pp. 109-113.

Wenzel, E. M. (1991). *Three-Dimensional Virtual Acoustic Displays*. NASA Technical Memorandum 103835. NASA Ames Research Center, Moffett Field, CA.

DESIGN CONSIDERATIONS FOR 3-D AUDITORY DISPLAYS IN COCKPITS

Robert H. Gilkey[1,2]
Brian D. Simpson[1]
Scott K. Isabelle[2,1]
Timothy A. Anderson[2]
Michael D. Good[3]

[1]Department of Psychology
Wright State University
Dayton, Ohio 45435 USA

[2]AL/CFBA
Wright-Patterson Air Force Base, Ohio 45433-7901 USA

[3]Honeywell Technology Center
Honeywell, Inc.
3660 Technology Drive
Minneapolis, Minnesota 55418 USA

SUMMARY

Potential cockpit applications of 3-dimensional auditory displays have generated considerable interest. These applications include: increasing speech intelligibility by spatially separating communication channels, providing a navigation beacon, directing pilots' attention to targets and threats, enhancing situational awareness by cuing a wingman's location or indicating an imminent collision, or even providing an auditory attitude indicator. However, cockpit noise and the complexity of the signals to be localized can adversely affect sound localization performance and may limit the effectiveness of these displays. We review the results of our experiments on sound localization in noise and the localization of speech signals with the head stationary, which indicate that although the ability to distinguish left from right can be quite accurate in adverse situations, often the accuracy of elevation judgments decreases and the number of front/back confusions increases with relatively small deviations from ideal conditions. The implications of these performance limitations for the design of auditory displays and potential strategies for enhancing performance will be discussed.

1. INTRODUCTION

The Duplex Theory (Rayleigh, Ref 1) identified interaural time differences and interaural level differences as the primary cues for sound localization. More than 100 years of research suggests that these overall interaural difference cues are sufficient to account for the ability of human listeners to determine the laterality of a sound, but are not sufficient to explain their ability to determine the elevation of sounds or their ability to determine whether sounds come from the front or from the rear.

More recent work by Blauert (Ref 2), Shaw (Ref 3), Oldfield and Parker (Ref 4, 5), Wightman and Kistler (Ref 6, 7), and others suggests that direction-dependent filtering of sounds by the torso, head, and pinna introduces "spectral cues" that allow the listener to determine both the elevation and front/back location of sounds. As a sound wave travels from a sound source to the eardrum, it is filtered because of the acoustic properties of the torso, head, and pinna. This filtering introduces direction-dependent peaks and

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

notches in the high-frequency (above 3-5 kHz) spectral envelope of the sound. It is believed that subjects make judgments about whether a sound is in front or in back, or above or below, based on recovering these direction-dependent changes in the spectral envelope. The importance of these spectral cues is also evident when sounds are presented through headphones. That is, when interaural differences are the only spatial cues available, a sound presented through headphones is localized as inside the head, rather than out in the environment. In contrast, when spectral cues are introduced, a "virtual" sound image is created that appears to be external to the head.

The availability of high-speed signal-processing hardware has made it possible to generate dynamic virtual sounds in real time by applying both interaural and spectral sound localization cues to stimuli (see for example McKinley, Ref 8, and Wenzel, Wightman, and Foster, Ref 9). Potential applications of this technology include virtual environments, telerobotics, architectural acoustics, home entertainment, and auditory displays. Here we focus on the potential applications of 3-dimensional auditory displays in cockpits. Such displays can help to increase communication effectiveness, maintain situational awareness, and aid flight control and targeting, by spatially separating communication channels, by indicating the locations of other aircraft, threats, and targets and by providing a navigation beacon or an attitude indicator.

The success of cockpit applications of 3-dimensional auditory displays will depend on the ability of pilots to judge easily, rapidly, and accurately the locations of the virtual signals. However, the effectiveness of current auditory displays is limited both by our incomplete knowledge of spatial hearing, and by design compromises introduced in order to ease implementation or circumvent processing limitations. Thus, users of auditory displays frequently report systematic misperceptions, including elevation errors and front-back confusions.

Here we address some of the limitations of our basic knowledge of spatial hearing. Most previous sound localization research has been performed in quiet environments using simple stimuli that are known to the subject a priori. In contrast, the cockpit environment is noisy, and the signals presented are often complex, carrying unknown semantic information in addition to location information. In this paper, we review the results of recent experiments from our laboratory on sound localization in noise and on the localization of speech stimuli, and discuss their relevance to the implementation of 3-D auditory displays in cockpits.

2. EXPERIMENT I - LOCALIZATION IN NOISE

Jacobson (Ref 10) and Perrott (Ref 11) evaluated the effect of interfering stimulation on spatial acuity (minimum audible angle, MAA) for narrowband targets. Both studies considered a relatively limited set of locations within the horizontal plane, and only considered spatial separations in azimuth. Nevertheless, the results of the two studies led to somewhat different conclusions: Jacobson suggested that noise had relatively little effect on spatial acuity even for sounds that were near masked threshold, whereas Perrott found large effects even for stimuli that were presumably well above threshold. The experiment by Good and Gilkey (Ref 12), reviewed here, provides a much more extensive examination of the effects of interference on sound localization performance. In their experiment, subjects made absolute location judgments for broadband targets in the presence of a broadband noise masker.

2.1. Method

The experiment was conducted in the Auditory Localization Facility of the Armstrong Laboratory at Wright-Patterson Air Force Base, Ohio. This facility includes a large anechoic chamber, which houses a 4.3-m diameter geodesic sphere with 277 loudspeakers mounted on its surface. The subjects sat with their heads in the middle of the sphere, and

held a bite-bar to limit head movements. A 20-cm diameter spherical model of auditory space was positioned in front of the subject at the waist level. After each stimulus presentation, the subject positioned the tip of an electromagnetic stylus at a point on the surface of the spherical model to indicate the perceived direction of the signal (for additional details, see Gilkey, Good, Ericson, Brinkman, and Stewart, Ref 13). A small visual display on the surface of the geodesic sphere in front of the subject provided information about trial timing.

The digitally generated signal was a train of 25-μs pulses, which repeated at a rate of 100 Hz; the train was windowed with 25-ms cosine-squared ramps to have a duration of 268 ms. The signal was bandpass-filtered from 0.53 kHz to 11.0 kHz. The digitally generated masker was a white Gaussian noise, windowed with 25-ms cosine-squared ramps to have a duration of 468 ms. It was bandpass-filtered from 0.41 kHz to 14.2 kHz. The overall level of the masker was approximately 52 dB SPL. Localization performance was examined for ten conditions: a quiet condition and nine masked conditions with signal-to-noise ratios ranging from +14 dB to -13 dB. (Signal-to-noise ratios were computed relative to the detection threshold for the case when the signal and masker were presented from the same speaker.)

On each trial, the signal location was randomly chosen from a set of 239 speaker locations that completely surrounded the subject in azimuth (360°) and ranged in elevation from -45° to +90°. The masker, when present, was always located at 0° azimuth and 0° elevation. During the course of the experiment, six localization judgments were measured at each of 239 speaker locations for each signal-to-noise ratio.

For ease of interpretation, we plot the data using the "3-pole" coordinate system. In this system, we consider the vector from the center of the subject's head to the actual target location or to the judged target location (the location vector). The left/right (L/R) coordinate is the angle between the location vector and the median plane (+90° indicating a location directly to the right of the subject, -90° indicating a location directly to the left of the subject), the front/back (F/B) coordinate is the angle between the location vector and the frontal plane (+90° indicating a location directly in front of the subject, -90° indicating a location directly behind the subject), and the up/down (U/D) coordinate is the angle between the location vector and the horizontal plane (+90° indicating a location directly above the subject, -90° indicating a location directly below the subject). This representation is useful in light of current views on sound localization, which suggest that the accuracy of judgments with respect to the L/R dimension is determined by the ability of the system to extract overall interaural time differences and interaural level differences, whereas performance in the F/B and U/D dimension is likely to depend on changes in the shape of the sound spectrum introduced by the pinna (e.g., see Kistler and Wightman, Ref 14). Because these cues are likely to be processed in quite different ways, we might expect an interfering stimulus to differentially disrupt the processing of these cues (e.g., a masker might interfere with the processing of spectral cues, but might not interfere with the processing of interaural difference cues).

2.2. Results and Discussion

Figure 1 shows selected results for a single subject. Each panel shows a scatter plot of the judged angle of the target as a function of actual angle for one dimension of the 3-pole coordinate system and a single signal-to-noise ratio. Each point is plotted at the center of a 5° x 5°-wide grid square. The size of each point indicates the proportion of judgments falling in that grid square out of the total number of judgments that could have fallen in that grid square. The top row of panels shows results for the L/R dimension, the middle row of panels shows results for the F/B dimension, and the bottom row of panels shows results for the U/D dimension. The left column of panels shows results obtained when the targets were presented in the quiet. Progressing to the right, each column shows results at a lower signal-to-noise ratio, with the far right column showing results for a signal-to-noise ratio of -10 dB. As

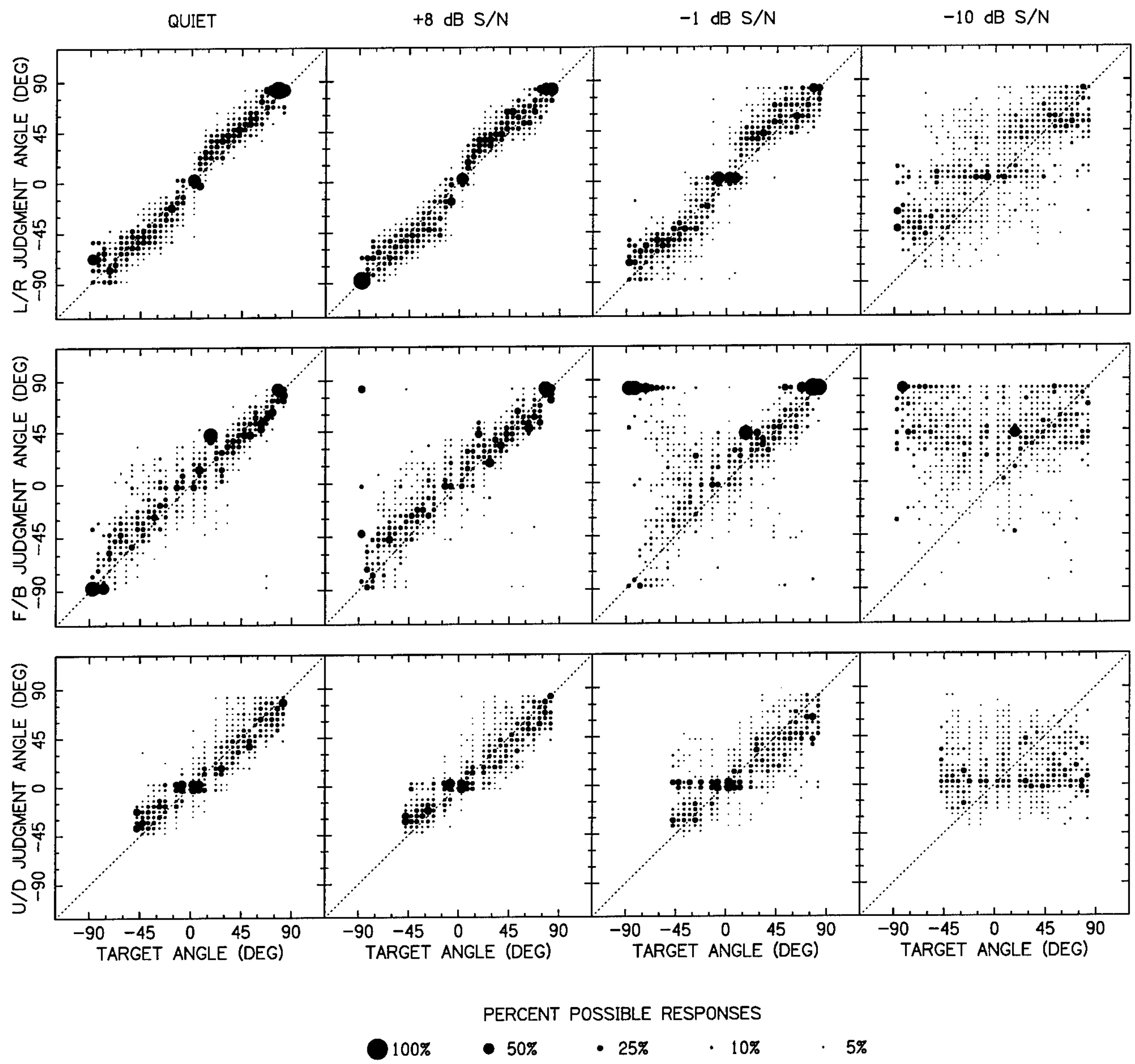


Figure 1. Scatter plots showing localization for subject MG in 3 dimensions at 4 different signal-to-noise ratios. See text for details. (After Good and Gilkey, Ref 12)

can be seen, performance in the quiet is very good for this subject in all 3 dimensions (i.e.,most points fall near the positive slope diagonal). As signal-to-noise ratio is lowered, the relation between the judged angle and the target angle weakens in all 3 dimensions. In the F/B dimension, there is no relation between the judged and target angle at a signal-to-noise ratio of -10 dB. Even at moderate signal-to-noise ratios, there are many front-back reversals (targets presented in the front hemi-sphere that are localized to the rear hemisphere, and vice versa), indicated by points falling in the upper-left or lower-right quadrant of the panel. At high signal-to-noise ratios, the effects of noise are less evident in the U/D dimension than in the F/B dimension, but at the lowest signal-to-noise ratios, there is again little relation between judged and target angles. In contrast, performance in the L/R dimension can be quite good at relatively low signal-to-noise ratios; even at the lowest signal-

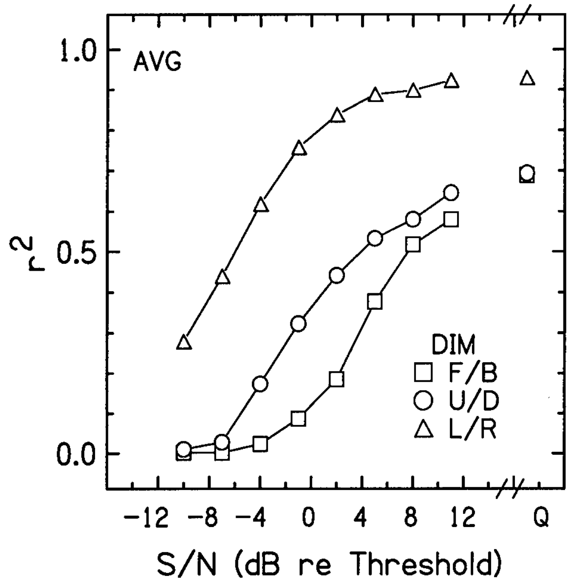


Figure 2. The proportion of variance accounted for by the relation between the judged and target angles averaged across 3 subjects is plotted as a function of signal-to-noise ratio for each of the 3 dimensions. (After Good and Gilkey, Ref 12)

to-noise ratio, some relation between judged and target angles is evident.

These results are summarized in Figure 2, which shows the proportion of variance accounted for by the relation between the judged and target angles, r^2. The average value of r^2 across the 3 subjects is plotted as a function of signal-to-noise ratio for each of the 3 dimensions. As can be seen, performance degrades nearly monotonically with decreases in signal-to-noise ratio in each dimension, but degrades much more quickly in the F/B and U/D dimensions than in the L/R dimension. The relation between judged and target angles in the L/R dimension accounts for nearly 30% of the variance, even at a signal-to-noise ratio of -10 dB. However, no relation between judged and target angles is observed for the F/B and U/D dimensions at the lowest signal-to-noise ratios. The decrease in r^2 for the F/B dimension corresponds to an increase in the number of back-to-front reversals; that is, at the lowest signal-to-noise ratio nearly all of the responses are biased toward the location of the masker such that essentially no responses occur in the rear hemisphere. However, because performance was examined for only a single masker location, it was difficult to be certain whether judgments were biased toward the masker or whether the subjects merely had a tendency to respond in the frontal hemisphere near the horizontal plane. The values of r^2 shown for the U/D dimension probably underestimate performance in that dimension because the range of actual target elevations was truncated (i.e., the range was -45° to +90° in the U/D dimension, and -90° to +90° in the L/R and F/B dimensions). Nevertheless, at the lowest signal-to-noise ratio, it is clear that the subjects receive essentially no information about the elevation of the sound.

It seems likely that the pattern of results observed in this experiment was at least partially dependent on the location of the masker. In particular, the fact that responses tended to be biased toward the masker (i.e., 0° L/R, 90° F/B, 0° U/D in this case) would most strongly effect performance for the dimension in which the masker was at an extreme position (i.e., the F/B dimension in this case). To address this concern, Good and Gilkey (Ref 15) considered the effects of changing the masker location on localization performance. If the poorer performance observed in the F/B dimension was due to the results of bias toward the masker, then one might expect better performance in the F/B dimension for some masker positions. On the other hand, if the acoustic cues that mediate performance in the F/B dimension are more susceptible to noise than those in the L/R or U/D dimensions, judgments in this dimension would be expected to remain less accurate than those in the other dimensions when the masker position is varied. On separate blocks of trials, the masker was presented from directly in front of the subject (0° L/R, 90° F/B, 0° U/D), directly to the left of the subject (-90° L/R, 0° F/B, 0° U/D), directly behind the subject (0° L/R, -90° F/B, 0° U/D), directly to the right of the subject (90° L/R, 0° F/B, 0° U/D), or directly above the subject (0° L/R, 0° F/B, 90° U/D). Two signal-to-noise ratios were individually chosen based on the subject's localization performance from the

previous study, performance in the L/R dimension determined the lower signal-to-noise ratio, and performance in the F/B and U/D dimensions determined the higher signal-to-noise ratio. The results reveal a complicated pattern of localization errors that are dependent on masker position and signal-to-noise ratio. In general, responses tend to be biased toward the location of the masker. However, this is not always the case. For example, at low signal-to-noise ratios with the masker to the side, responses tend to be biased away from the midline, some toward the masker and some away from the masker. Overall the pattern of results observed by Good and Gilkey (Ref 12) was also observed in this experiment. That is, although the single masker location they examined (in front) may have had its strongest effects on the F/B dimension, performance in the F/B dimension is generally poor and easily disrupted by noise for all masker locations. Similarly, performance in the L/R dimension is good in general and is robust to the influence of noise.

These results suggest that auditory displays are likely to provide relatively good information in the L/R dimension, even in adverse acoustic environments. In contrast, F/B judgments, and to a lesser degree U/D judgments, may be severely disrupted when noise is present.

3. EXPERIMENT II - LOCALIZATION OF SPEECH STIMULI

Of the sounds present in a cockpit, speech communication signals are of central importance. Although the primary role of these signals is to deliver semantic information to the pilot, they could also carry important spatial information. For example, spatial cues could be added to a wingman's communication channel, such that the wingman's voice would appear to come from the direction of his or her plane. Such spatial information could significantly enhance situational awareness. However, for this information to be useful, the auditory system must be able to separate this spatial information from the semantic information already present in the signal. If the ability of subjects to determine the L/R coordinate is dependent on low-frequency interaural time differences, as suggested by Wightman and Kistler (Ref 16), then because speech contains significant low-frequency energy we would expect it to be accurately localized in the L/R dimension. On the other hand, information about the F/B dimension and the U/D dimension is believed to be coded via direction-specific changes in the high-frequency spectrum of the sound. Because speech has less high-frequency energy, it may be less effective at carrying information about the F/B and U/D dimensions. Moreover, because the spectral envelope of speech is irregular and varies from utterance to utterance, the auditory system may find it difficult to distinguish between those spectral modulations introduced by the filtering of the torso, head, and pinna, and those spectral modulations inherent in the speech stimulus. Recovering the spectral variations introduced by the torso, head, and pinna is straightforward when the sound source spectrum is flat (i.e., any peaks and notches present in the spectrum of the sound received at the tympanic membrane must have been introduced by the filtering of the torso, head, and pinna). However, when the sound source spectrum is irregular, the auditory system may have difficulty separating the spectral variations introduced by the torso, head, and pinna from those that were present in the original sound source. That is, speech may not serve as an effective carrier waveform for F/B and U/D spatial information.

The few studies that have examined localization for speech have only considered sound-source locations that varied in azimuth within the horizontal plane (Ericson, McKinley, and Valencia , Ref 17; Begault and Wenzel, Ref 18; Ricard and Meirs, Ref 19; Shigeno and Oyama, Ref 20; Koehnke, Besing, Goulet, Allard, and Zurek, Ref 21). Most of these studies found that performance in the L/R dimension was comparable for speech and non-speech stimuli. Some studies found that F/B confusions increased with speech stimuli, while others found similar levels of F/B confusions for speech and non-speech stimuli. Although the elevation of the sound source was not varied in any of these studies, Begault and Wenzel

required the subjects to make elevation judgments; these judgments were less accurate for speech than for non-speech signals. In the experiment of Gilkey and Anderson (Ref 22), reviewed here, the localization of speech and non-speech stimuli is compared in both azimuth and elevation.

3.1 Method

This experiment was also conducted in the Auditory Localization Facility; the apparatus and response collection procedure were the same as those described in Experiment I. The speech targets were 266 words from the Modified Rhyme Test (MRT), read by both a male and a female talker, and digitized into a personal computer, where they were scaled to make the average levels of the male and female speech approximately equal. After scaling, the 532 individual speech tokens varied in level over an approximate 17-dB range and varied in duration from approximately 215 ms to approximately 750 ms. The click target was a train of 25-µs pulses, which repeated at a 100-Hz rate, and was windowed with 25-ms linear ramps to have a duration of 466-ms. The clicks were scaled to have approximately the same level as the average level of the speech. The signals were presented at an average level of approximately 53 dB SPL.

On each trial, the signal location was randomly chosen from the same pool of 239 possible speaker locations that surround the subject in azimuth (360°) and ranged from -45° to 90°. When speech targets were used, both the talker and word were randomly chosen from trial to trial. During the course of the experiment, 5 localization judgments were measured at each of the 239 speaker locations for the speech targets and for the click targets.

3.2. Results and Discussion

Figure 3 plots the difference in RMS error between the speech target and click target conditions for each of 4 subjects in each of the 3 dimensions of the 3-pole coordinate system. Each cluster of bars shows results for one dimension of the 3-pole coordinate system.

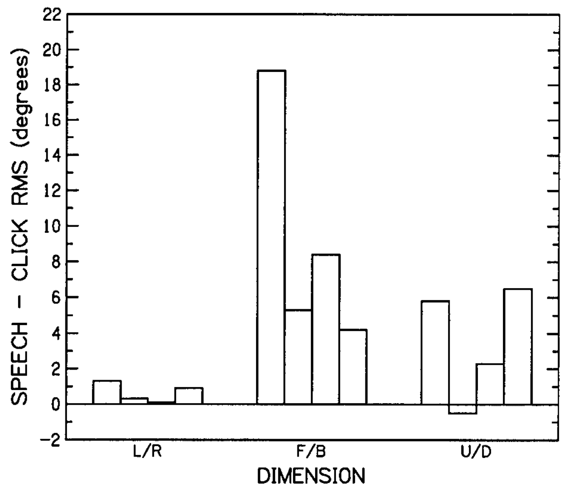


Figure 3. The difference in RMS error between the speech target and click target conditions for each of 4 subjects in each of the 3 dimensions of the 3-pole coordinate system is plotted. Each cluster of bars shows results for one dimension of the 3-pole coordinate system. Within each cluster, each bar shows the results for a different subject. That is, the left-most bar in each cluster shows the results for subject 1, the right-most bar shows results for subject 4, etc. (After Gilkey and Anderson, Ref 22)

Within each cluster, each bar shows the results for a different subject. In each dimension, we compute the RMS error between the judgment angle and the target angle. As can be seen, in agreement with previous studies on speech localization, performance in the L/R dimension is similar for speech and non-speech stimuli. However, for all subjects, performance in the F/B dimension is worse for speech than for nonspeech. Similarly, for all but one subject, performance in the U/D dimension is worse for speech than for non-speech.

These results are compatible with the expectation stated at the beginning of Section 3 that speech is effective at carrying information about the L/R dimension. Presumably, the low-frequency energy in speech allows for an effective representation of interaural differences. On the other hand, speech is relatively ineffective at carrying information about the F/B and U/D dimension. Presumably, because speech has comparatively

little high-frequency energy, and because the spectral envelope of speech is irregular and varies from utterance to utterance, it is difficult for the auditory system to recover the direction-dependent spectral signature of the torso, head, and pinna. The designers of 3-D auditory displays are likely to apply spatial information to speech communication signals. In a cockpit, this strategy has the advantage of supplying spatial information without adding to the total number of auditory signals. However, designers need to consider the effectiveness of speech as a carrier and the performance implications of possible localization accuracies in the F/B and U/D dimensions.

4. GENERAL DISCUSSION

The results of these two studies indicate that previous laboratory studies of sound localization may not be representative of the performance that can be expected with auditory displays in applied settings. Specifically, sound localization accuracy is reduced when noise is present. In addition, sound localization accuracy is reduced when the targets are speech sounds. In both cases, performance in the L/R dimension can be quite good, similar to that observed for spectrally flat stimuli in the quiet. In contrast, performance in the F/B and U/D dimensions, appears to be more easily disrupted.

It is important to note that the impact of these results on pilot effectiveness in applied settings remains to be determined. In both studies reported here, head movements were restricted. In contrast, pilots will be free to move their heads. Head movements may mitigate against these effects by providing additional dynamic sound localization cues, or by altering the effectiveness of static sound localization cues when the head is reoriented in the sound field. In addition, a different pattern of errors may emerge when the interfering noise is less directional than the maskers employed in our experiments. In Experiment II, we used isolated words as targets, it is possible that longer utterances might lead to better performance. If, however, the results reported here hold in applied settings, then the designers of 3-D auditory displays should consider alternate ways to represent F/B and U/D spatial information. For example, if accurate localization in the F/B dimension is critical, the same signal processing hardware that is used to introduce spatial cues could be used to change the pitch or the timbre of a target when it is presented in the rear hemisphere. Better localization performance for speech stimuli might be achieved by enhancing or adding high-frequency information.

It should also be noted that, even in situations where localization accuracy is reduced, 3-D auditory displays may provide considerable benefit. For example, one of the promising uses of 3-D auditory displays is to direct the attention of the user toward relevant information. For example, spatialized sound could be used to direct a pilot's attention to an important visually displayed instrument, or outside the cockpit to a potential threat. Previous research in the Armstrong Laboratory by Perrott, Cisneros, McKinley, and D'Angelo (Ref 23) has shown that search times for an isolated light against a dark background can be reduced by 10-50% when an auditory cue was present. In a recent pilot experiment conducted in our laboratory, using a more difficult visual search task, we found a much larger effect of the auditory cue.

In our experiment, the subject wore a head-mounted display (HMD) with a limited field-of-view (40° horizontal by 20° vertical), and looked at an array of virtual letters that surrounded him/her in azimuth, and ranged from -30° elevation to +30° elevation. All of the letters, except the target, were either capital "Ps" or capital "Qs." The subject's task was to find the single capital "R" (i.e., the target). Characters were positioned in 5° x 4° grid cells, such that approximately 40 characters were visible in the HMD at any time. The subject searched the entire field of letters until the "R" was found. In the auditory-aided condition, a virtual auditory cue (filtered with head-related transfer functions, presented through headphones, and fixed in virtual space using a head tracker) was presented near the virtual spatial location of the visual target.

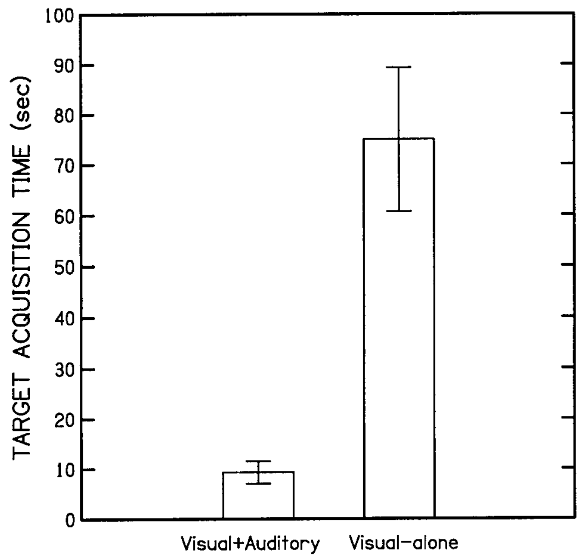


Figure 4. Target acquisition times, averaged across 5 subjects, are plotted for the visual-only and auditory-aided visual search conditions.

Figure 4 shows target acquisition times averaged across 5 subjects for the visual-only and auditory-aided visual search conditions. Acquisition times decreased by more than a factor of 8 when the auditory cue was added. Note also that this increase in speed was realized with a relatively poor auditory display (e.g., non-individualized head-related transfer functions, no reverberation model, and no interpolation between recorded spatial locations such that the auditory signal could be as much as 9° away from the center of the visual target). Thus, even if the localization inaccuracies we observed in the F/B and U/D dimensions are also observed in applied settings, it seems likely that auditory displays can still provide substantial benefits.

5. CONCLUSIONS

Auditory displays have considerable potential for enhancing mission effectiveness. The results presented here suggest that localization accuracy in the F/B and U/D dimensions may be degraded in many applied settings relative to that measured in the laboratory. The designers of 3-D auditory displays should consider the impact of these reductions in localization accuracy on their particular application and, if necessary, find mechanisms to augment localization performance.

6. REFERENCES

1. Rayleigh, Lord (J.W. Strutt, 3rd Baron of Rayleigh), "On our perception of sound direction," Philos. Mag., *13*, 1907, pp 214-232.

2. Blauert. J., "Spatial Hearing," Cambridge, MA: MIT Press, 1983.

3. Shaw, E.A.G., "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am., *56*, 1974, pp 1848-1861.

4. Oldfield, S.R., and Parker, S.P.A., "Acuity of sound localization: A topography of auditory space. I. Normal hearing conditions," Perception, *13*, 1984, pp 581-600.

5. Oldfield, S.R., and Parker, S.P.A., "Acuity of sound localization: A topography of auditory space. II. Pinna cues absent," Perception, *13*, 1984, pp 601-617.

6. Wightman, F.L., and Kistler, D.J., "Headphone simulation of free-field listening. I: Stimulus synthesis," J. Acoust. Soc. Am., *85*, 1989, pp 858-867.

7. Wightman, F.L., and Kistler, D.J., "Headphone simulation of field listening. II: Psychophysical validation," J. Acoust. Soc. Am., *85*, 1989, pp 868-878.

8. McKinley, R. L., "Concept and Design of an Auditory Localization Cue Synthesizer," Master's thesis, AFIT/GE/ENG/88D-29, 1988.

9. Wenzel, E.M., Wightman, F.L., and Foster, S.H., "A virtual display system for conveying three-dimensional acoustic information," Proc. Hum. Fac. Soc., *32*, 1988, pp 86-90.

10. Jacobsen, T., "Localization in noise," (Tech. Rep. No. 10). Denmark: Technical University of Denmark Acoustics Laboratory, 1976.

11. Perrott, D.R., "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity," J. Acoust. Soc. Am., *75,* 1984, pp 1201-1206.

12. Good, M.D., and Gilkey, R.H., "Sound localization in noise: Effects of signal-to-noise ratio," J. Acoust. Soc. Am., *99*, 1996, pp 1108-1117.

13. Gilkey, R.H., Good, M.D., Ericson, M.A., Brinkman, J, Stewart, J.M., "A pointing technique for rapidly collecting localization responses in auditory research," Behavior Res. Methods Instr. Comput., *27*, 1995, pp 1-11.

14. Kistler, D.J., and Wightman, F.L., "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," J. Acoust. Soc. Am., *91*, 1992, pp 1637-1647.

15. Good, M.D., and Gilkey, R.H., "Auditory localization in noise: II. Effects of masker location," J. Acoust. Soc. Am, *95*, 1994, pp 2896.

16. Wightman, F.L, and Kistler, D.J., "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am., *91*, 1992, pp 1648-1661.

17. Ericson, M.A., McKinley R.L., Valencia G., "Human auditory localization performance in azimuth," Tech. Rep. No. AL-TR-1990-0080, Wright-Patterson Air Force Base, Armstrong Laboratory, 1991.

18. Begault, D.R., and Wenzel, E.M., "Headphone localization of speech," Hum. Fac., *35* , 1993, pp 361-376.

19. Ricard G.L, and Meirs S.L., "Intelligibility and localization of speech from virtual directions," Hum. Fac., *36*, 1994, pp 120-128.

20. Shigeno, S, and Oyama, T., "Localization of speech and non-speech sounds," J. Psychol Res., *25*, 1983, pp 112-117.

21. Koehnke, J., Besing, J., Goulet, C., Allard, M., and Zurek, P.M., "Speech intelligibility, localization, and binaural detection with monaural and binaural amplification," J. Acoust. Soc. Am., *94*, 1992, pp 24-34.

22. Gilkey, R.H., and Anderson, T.R., "The accuracy of absolute localization judgments for speech stimuli," J. Vestib. Res., *5*, 1995, pp 487-497.

23. Perrott, D.R., Cisneros, J., McKinley, R.L., and D'Angelo, W.R., "Aurally aided detection and identification of visual targets," Proc. Hum. Fac. Soc., *39*, 1995, pp 103-108.

Synergie perceptuelle et cognitive dans l'orientation vers une cible : SON 3D et information visuelle

M. COURNEAU, F. LEPPERT, C. GULLI, A.LEGER- SEXTANT AVIONIQUE - BP 91 - Rue Toussaint-Catros - 33 166 Saint-Médard-en-Jalles - France

L. PELLIEUX - IMASSA /CERMA -BP 73 - 91223 Brétigny sur Orge Cedex - FRANCE

M. HAAS - ARMSTRONG LABORATORY - AL CFHP - 2255 H Street - Wright-Patterson AFB, OH, 45433-7022 - U.S.A.

RESUME

Dans l'expérimentation décrite ici, deux types d'information pouvant permettre la localisation d'une menace ont été étudiés dans un contexte aéronautique. Une information visuelle, consistant en une flèche pointant vers la menace, faisant appel à un mécanisme cognitif de relativement haut niveau, était présentée sur un HMD. Une aide à la localisation plus perceptuelle, basée sur le son 3D, était utilisée en alternance ou simultanément. L'étude a été menée dans le cadre d'une coopération entre Sextant Avionique, Armstrong Laboratory et l'IMASSA / CERMA. Son but était d'évaluer l'efficacité de ces deux modalités d'information dans un simulateur de vol et de tester l'hypothèse d'une synergie entre elles. Les résultats présentés ici s'attachent plus particulièrement à la phase d'orientation vers la menace. L'analyse des données recueillies pendant l'expérimentation montre que les informations visuelles et sonores sont équivalentes et qu'il existe bien une synergie additive. Cette synergie est révélée par une amélioration significative des performances des sujets lorsque les deux modalités sont présentées d'une manière additive et simultanée.

MOTS-CLES

Son 3D - Simulation de vol - Information sonore - Information visuelle - Synergie perceptuel / cognitif

TABLE DES MATIERES

1. INTRODUCTION

La localisation visuelle et l'identification d'un élément de l'environnement extérieur est dans de nombreux cas une tâche critique pour le pilote d'avion d'armes en particulier, mais aussi pour l'aviation commerciale. Il peut s'agir d'une menace, d'une cible, d'un autre avion à éviter... Il est donc primordial de présenter au sujet une ou des informations permettant une localisation suffisamment précise et rapide de la source de menace.

Dans le domaine de l'aviation commerciale, plusieurs auteurs ont déjà tenté d'évaluer l'apport de l'information sonore localisée dans les évitements de collision (1). Leurs résultats ont montré, avec des informations visuelles présentées en tête basse, que l'information contenue dans le son 3D permettait une diminution sensible des temps de localisation des menaces par rapport à diverses situations de référence. Plus récemment, Bronckorst (2) a effectué un travail similaire en simulateur d'avion de combat sans cependant utiliser une symbologie visuelle asservie aux mouvements de la tête.

Dans le cadre du développement de symbologie de viseur de casque, SEXTANT Avionique a été conduit à étudier les aides à la localisation utilisant les potentialités de ces nouveaux équipements. Parallèlement, en collaboration avec l'IMASSA/CERMA, des études sur le son 3D ont également été menées pour identifier les domaines de précision accessibles en particulier avec des fonctions de transfert personnalisées. Des travaux similaires se déroulant à l'Armstrong Laboratory, une nouvelle expérimentation a été conduite dans le cadre du protocole d'accord Franco-Américain "Supercockpit". Elle avait pour objectif principal l'évaluation de l'efficacité du couplage d'une symbologie visuelle, présentée dans un viseur de casque, avec l'information sonore tridimensionnelle dans une tâche de localisation rapide de menace. Le but de l'expérimentation consistait en particulier à tester l'hypothèse d'une synergie additive entre les deux modalités d'aide lorsqu'elles sont présentées simultanément.

2. METHODOLOGIE

2.1. Sujets

Douze sujets ont participé à l'expérimentation. Cette population, essentiellement composée de pilotes, demeure relativement peu homogène quant à l'expertise de chacun : pilotes d'essais, pilotes de chasse en activité ou non, pilotes expérimentateurs et expérimentateurs navigant d'essais, voire pilote privé. Tous avaient en commun, avec des degrés divers, une maitrise du pilotage de la plateforme adéquate pour les buts de l'expérimentation. En revanche certains pilotes avaient une expertise préalable avec les

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

visuels de casque, alors que d'autres découvraient ce moyen nouveau. De même, l'expérience du son 3D était inégalement répartie dans la population.

2.2. Dispositif Experimental

2.2.1. Le casque

L'un des éléments les plus importants de cette expérimentation est le casque du pilote, dit "DE Grand Champ", développé par Sextant Avionique. Ce casque a les deux fonctions suivantes :

- la projection sur sa visière d'une image vidéo : le casque est binoculaire, avec un recouvrement de 50 %, et présente un champ total de 40° x 30°.
- la détection de la position et de l'orientation de la tête du pilote, grâce à un capteur intégré au casque et à un émetteur placé dans la cabine de simulation. Cette D.D.P. électromagnétique a été développée et réalisée par Sextant Avionique. C'est un élément fondamental de l'expérimentation, car elle permet, comme on le verra plus loin, d'asservir l'image du simulateur ainsi que le son aux mouvements de tête du pilote.

De plus, il a été muni d'écouteurs stéréophoniques de haute qualité, supérieure aux écouteurs standards monophoniques.

2.2.2. Le cockpit

Le sujet est installé dans la maquette d'un cockpit d'avion d'armes, comportant outre le système de Détection De Position, un manche et une manette des gaz latéraux, permettant le pilotage, et un amplificateur audio permettant le réglage du volume du son arrivant dans les écouteurs du casque.

2.2.3. Le simulateur de vol

Le paysage extérieur, ainsi que les symbologies de pilotage, sont calculés par un simulateur de vol, développé sur Silicon Graphics.

L'image générée dépend :

- d'une base de données géographique simplifiée
- des manoeuvres de la plateforme commandées par le pilote
- de l'orientation de la tête du pilote dans la cabine.

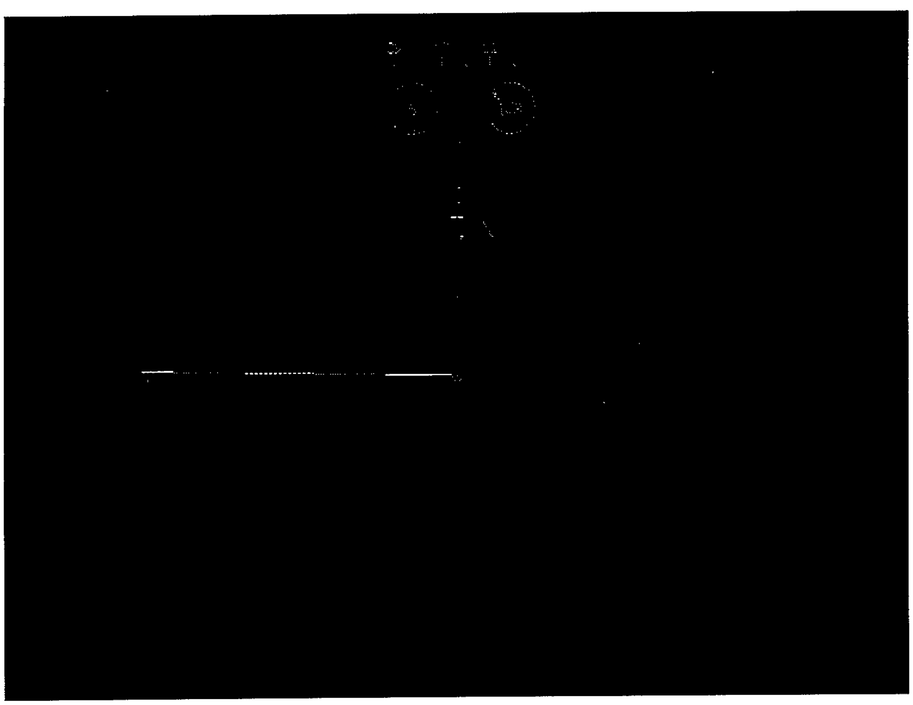
Elle est projetée sur la visière du casque du sujet.

2.2.4. Information Visuelle

Outre le paysage synthétique et une symbologie de pilotage HUD (Head-Up Display) dérivée du mode de navigation du Mirage 2000, une symbologie HMD (Helmet Mounted Display) est présentée sur la visière du casque, développée par Sextant Avionique au cours d'expérimentation préalable (3).

Cette symbologie apparaît lorsque le pilote regarde en dehors du HUD, et permet de connaître l'attitude, la pente, la vitesse et l'altitude de l'avion.

De plus, le guidage vers le prochain but de navigation est assuré par une flèche asservie aux mouvements de tête du pilote, dont la direction pointe vers ce but et de longueur fonction de la distance angulaire entre la ligne de visée du pilote et le but. La flèche disparaît lorsque le but entre dans le champ du vision du casque.



Symbologie APS (Aide à la Perception de la Situation) présentée dans le viseur de casque.

Pour notre expérimentation, cette flèche a été utilisée comme indicateur de menace. En résumé, cette flèche guide toujours vers le but premier de la mission : but de navigation en phase normale, menace à partir du moment où elle est déclenchée jusqu'à son acquisition par visée et validation appui.

Ce mécanisme de guidage fait ainsi appel à une modalité visuelle fortement cognitive dans la mesure où la direction et la longueur du symbole (métrique non perceptive) guide la direction et l'amplitude des mouvements de tête .

2.2.5. *Information Sonore*

Le son 3D permet de générer une image sonore telle que l'auditeur perçoive les sons comme issus d'un point particulier de l'espace.

Une précédente étude menée avec l'IMASSA/CERMA a permis la réalisation d'un outillage acoustique et d'une technique de mesure et calcul des fonctions de transfert des sujets inspirée de celles utilisées par Pösselt et Whigthman(5, 8).

Une expérimentation sur la précision de localisation d'une source avait alors été réalisée : elle consistait à localiser une source sonore sans référence visuelle, le plus précisément possible, sans consigne de vitesse. L'information sonore tridimensionnelle était diffusée jusqu'à désignation de la source, et l'orientation de la tête du sujet était prise en compte pendant la recherche (4).

Cette étude a conduit aux conclusions suivantes :

- la précision de localisation est meilleure avec des fonctions de transferts personnalisées.

  Ces fonctions permettent de reconstituer les effets du filtrage acoustique réalisé par le thorax, la tête et les pavillons du sujet, qui entrent en jeu dans le mécanisme de localisation d'un son (le filtrage réalisé par les pavillons joue en particulier un rôle sensible dans la localisation en site).
- pour les meilleurs sujets, l'erreur RMS de localisation est de 4° en gisement, 5° à 7° en site, 6 à 8° sur la ligne de visée.
- un apprentissage, plus ou moins poussé suivant les individus, améliore les performances

Le SON 3D, dont l'efficacité a ainsi été démontrée, semble intéressant pour transmettre l'information de localisation recherchée ; en effet, cette modalité fait appel essentiellement à un mécanisme perceptuel profondément enraciné dans le comportement de l'homme mais aussi de nombreuses espèces animales. Cependant, bien que la nature de l'information la place essentiellement dans la sphère perceptive, des mécanismes cognitifs sont également mis en jeu dans la représentation mentale de la situation, comme en témoigne l'amélioration de performance liée à l'entraînement.

2.3. Protocole Experimental

Il était alors intéressant d'évaluer l'efficacité de cette information lorsque le sujet a une charge de travail non nulle, dans un domaine opérationnel suffisamment réaliste.

Lors de cette nouvelle évaluation, les sujets doivent tout d'abord accomplir du mieux possible une tâche primaire, consistant à piloter un simulateur de vol et suivre des buts de navigation, tout en respectant des consignes de vitesse et d'altitude.

A un moment donné, aléatoire, une alarme se déclenche ; la mission première du pilote devient alors la localisation visuelle la plus rapide possible de la menace. Toute symbologie de guidage vers le but de navigation est rendue inactive, cette tâche étant alors devenue secondaire. La menace, fixe dans l'espace, est matérialisée par une sphère blanche, positionnée en limite de portée visuelle pour la résolution du simulateur, soit à une distance d'environ 8 Milles Nautiques. Pour valider sa localisation, le pilote doit la viser avec le réticule de désignation lié au casque avec une précision inférieure à 3.5° et appuyer alors sur la gâchette de tir. A ce moment, si la visée est correcte, un cercle apparaît autour de la menace pendant 3 secondes, et les informations visuelles ou / et sonores d'aide à la localisation disparaissent.

Une fois cette mission accomplie, le pilote revient à sa tâche de navigation.

Dans la suite du document, l'aide sonore à la localisation sera désignée comme "LAW" ("Localized Audio Warning") et l'aide visuelle en HMD sera désignée comme "TDC" ("Target Designation Cue").

Afin d'évaluer les deux types d'informations présentés et leur combinaison, quatre configurations ont été testées :

- <u>NLAW / NTDC</u> : (No LAW / No TDC) : cette configuration ne comporte ni aide visuelle en HMD, ni aide sonore tridimensionnelle. Lorsque la menace se déclenche, une alarme sonore ne contenant aucune information de localisation se fait entendre, et la menace apparaît sur une visualisation de VCM (Visualisation Contre-Mesure), donnant ainsi la position de la menace par rapport à l'avion en deux dimensions. C'est la configuration de référence, correspondant à la situation rencontrée actuellement sur un avion d'armes.
- <u>NLAW / TDC</u> : ici, seule l'information visuelle en HMD est présentée, une alarme sonore non localisée se déclenchant au moment de l'apparition de la menace.
- <u>LAW / NTDC</u> : ici, seul le SON 3D intervient comme aide à la localisation de la menace.
- <u>LAW / TDC</u> : cette configuration fait intervenir les deux modalités d'une manière simultanée.

2.4. Déroulement de l'expérimentation

L'expérimentation a commencé, pour chacun des 12 pilotes y participant, par l'acquisition de leurs fonctions de transfert, grâce aux outils développés au cours de l'étude ultérieure.

Une première précaution a été prise en faisant subir aux sujets, conformément aux conclusions de cette précédente étude, un premier apprentissage au SON 3D, afin d'une part de les habituer à ce type d'aide à la localisation, d'autre part de vérifier que leurs fonctions de transfert ne comportaient pas de défaut.

Puis, un second apprentissage, plus poussé, leur a permis dans un premier temps de se familiariser avec le pilotage du simulateur et les symbologies présentées, tout particulièrement en HMD, puis à la mission proprement dite, avec localisation de menace avec les deux types d'aides, visuelle et sonore.

Enfin, l'évaluation elle-même a consisté, pour chacun des 12 pilotes, à la réalisation de 27 scénarios (9 cibles présentées 3 fois dans un ordre aléatoire) pour chacune des quatre configurations. Ces quatre sessions ont été présentées aux pilotes dans un ordre suivant un carré latin.

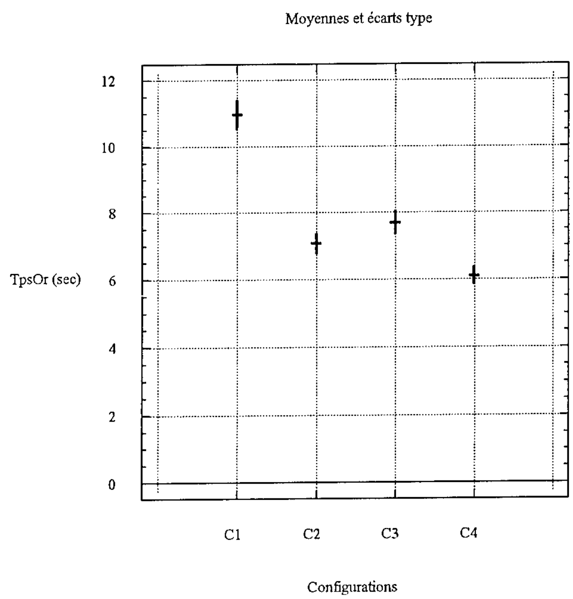
3. RESULTATS

3.1. Statistiques sur les temps d'orientation

Nous nous attachons ici à la phase d'orientation, qui s'étend du déclenchement de l'alarme jusqu'à ce que la direction de la menace entre dans le champ du visuel de casque.

Statistiques sur les temps d'orientation (sec.)

	Global	C1	C2	C3	C4
Moyenne	7.96	**10.96**	**7.07**	**7.68**	**6.08**
s	5.82	**7.13**	**4.76**	**5.62**	**4.15**
Min.	0.27	0.27	0.98	0.82	0.77
Max.	50.15	50.15	30.09	31.15	20.38



où C1 = NLAW / NTDC
C2 = NLAW / TDC
C3 = LAW / NTDC
C4 = LAW / TDC

L'étude des moyennes et écarts type des temps d'orientation montre que la configuration de référence se détache nettement des autres, avec des temps d'orientation vers la menace bien plus importants.

Les aides à la localisation visuelle et sonore, donnent des résultats pratiquement équivalents, l'information visuelle ayant une courte avance en terme de moyenne et surtout un écart type plus réduit.

Enfin, la configuration "additive", présentant les deux modalités simultanément, donne nettement les meilleures performances, tant sur la moyenne que sur l'écart type de la durée de la phase d'orientation.

3.2. Analyse de la variance

L'analyse de la variance aboutit à la représentation des groupes homogènes de configuration suivante :

Configurations	Moyennes	Groupes
LAW / TDC	6.08	X
NLAW / TDC	7.07	X
LAW / NTDC	7.68	X
NLAW / NTDC	10.96	X

Elle confirme les résultats précédents et nous montre que :

- les informations présentées, tant visuelle en HMD que sonore tridimensionnelle, permettent une amélioration significative (p<0.001) par rapport à la configuration de référence ne donnant qu'une information 2D à contenu fortement cognitif et non asservie aux mouvements de tête.
- les modalités visuelle et sonore utilisées ne présentent pas de différence significative.
- la synergie des deux modalités d'information de localisation (visuelle et sonore) est vérifiée, puisque l'amélioration obtenue grâce à la configuration "additive" est statistiquement significative (p<0.001).

4. DISCUSSION

Outre les différentes configurations de présentation d'information utilisées lors de l'expérimentation, de nombreux facteurs sont susceptibles d'influencer la performance du pilote dans ce type de tâche.

L'analyse de la variance montre, pour la phase d'orientation vers la menace, une très forte influence de l'écart angulaire de la cible par rapport à l'avion au déclenchement de la menace, le gisement ayant un rôle prépondérant : cela n'est pas surprenant, puisqu'il faut tourner la tête davantage, voire manoeuvrer l'avion lorsque la cible est trop loin angulairement.

Il existe de plus, en dépit des précautions prises (apprentissage préalable au son 3D et au simulateur), un effet significatif "ordre de test". Cet effet a parfois été très important, en particulier avec un des pilotes dont les résultats du premier test ont été relativement catastrophiques. Ces résultats ont néanmoins été inclus dans l'analyse, dans la mesure où le sens des variations observé était cohérent avec l'ensemble des résultats.

D'une manière générale, l'utilisation du protocole d'expérience "carré latin" permet de neutraliser les effets d'ordre pour l'analyse statistique des résultats. La technique utilisée est d'autre part connue pour sa grande robustesse, ce qui fait que l'on peut être confiant sur la pertinence des résultats obtenus.

La population de sujets ayant participé à l'expérimentation est relativement non homogène : elle recouvre en effet une assez large variété d'expertises (pilotes d'essais en activité, pilotes de chasse en activité ou non, pilotes "expérimentateurs", voire simple pilote privé). Parmi ces différentes expertises, certains sujets avaient déjà participé à des expérimentations avec viseur de casque, d'autres découvraient pour la première fois ce moyen de simulation nouveau. De même, l'expérience du son 3D était inégalement répartie dans la population.

Comme dans la plupart des expérimentations sur la performance humaine, il a été constaté un "effet sujet" fortement significatif, avec des valeurs extrêmes parfois importantes. Il est cependant intéressant de remarquer que ces écarts ne semblent pas systématiquement corrélés avec l'expertise des différents individus. Les différences relèvent sans doute plus d'aptitudes individuelles, en particulier en ce qui concerne le son 3D, que d'une expertise préalable.

On note toutefois que, quelle que soit l'expertise du sujet, l'ordre de grandeur de la performance et des gains obtenus par rapport à la situation de référence demeure globalement assez constant. Ceci souligne le caractère "perceptuel" et "intuitif" de l'aide apportée par le son 3D lors d'une tâche, que l'on doit cependant qualifier d'élémentaire.

Cela confirme aussi, comme généralement trouvé, que les performances dans la réalisation d'une tâche sont accrues lorsque des informations similaires sont présentées à travers plusieurs modalités plutôt qu'une seule (6). Par rapport aux expérimentations antérieures (1, 2), les résultats obtenus apparaissent cohérents, certaines différences entre les résulats obtenus avec le son 3D et la symbologie visuelle étant vraisemblablement liées à l'utilisation d'une symbologie d'aide visuelle tenant compte de l'orientation de la tête.

En revanche, dans la tâche essentiellement visuelle de localisation fine de la menace, aucune synergie n'a été observée, la condition utilisant le son 3D ne se différenciant pas de la situation de référence. La nature de l'expérimentation ne permet pas de statuer sur la compatibilité des deux symbologies dans cette phase, car aussi bien une exclusion mutuelle liée à la tâche que les caractéristiques de la symbologie sonore peuvent expliquer ce résultat.

Le point essentiel, sur le plan fondamental, est la mise en évidence d'une synergie entre une information visuelle à fort contenu cognitif et une information auditive de nature plus perceptive, ceci dans la phase d'orientation vers la menace. Ceci montre la compatibilité des sources d'information utilisées et celle des ressources utilisées pour leur traitement. On retrouve ici une analogie avec les théories de compatibilité stimulus-réponse proposées par Wickens (7). Les deux informations ont un caractère redondant, comme l'indique l'analyse de variance, mais se révèlent également complémentaires pour la phase d'orientation vers la menace. Il est difficile de savoir s'il s'agit d'une réelle complémentarité, avec l'hypothèse d'une représentation précise et instantanée de l'écart angulaire fournie par le son 3D, ou bien d'une potentialisation mutuelle plus globale des deux informations. L'augmentation du niveau de confiance dans le modèle de localisation (image mentale) pourrait alors expliquer les résultats. En tout état de cause, lorsque le signal sonore ne porte plus d'information (recherche visuelle fine), il est purement et simplement occulté sans interférence négative avec la symbologie visuelle, ainsi que l'indique encore une fois l'analyse statistique.

L'aide apportée par le son 3D apparaît donc de nature très intuitive, ce qui paraît a priori une bonne chose dans un environnement d'avion de combat. Il faut signaler encore une fois ici que la simplicité de l'environnement simulé et des tâches effectuées ne permet pas de conclure formellement sur d'éventuels problèmes d'intégration de ce type d'aide dans un environnement complexe impliquant des processus cognitifs de haut niveau. Cette expérimentation a le mérite de montrer l'efficacité du son 3D dans un processus d'orientation vers une cible et sa complémentarité avec les aides visuelles plus classiques. La faisabilité et l'intérêt de ce concept apparaissent cependant assez clairement démontrés, ouvrant ainsi la porte à des recherches ultérieures précisant son intérêt opérationnel dans un environnement plus représentatif.

5. CONCLUSION

La présente étude a montré que le son 3D constituait une source d'information au moins équivalente à une aide cognitive présentée sous forme de symbologie visuelle. L'aide à la localisation spatiale apportée par le son 3D présente un caractère perceptif direct. Une synergie additive de cette symbologie sonore avec une symbologie visuelle essentiellement cognitive a également été mise en évidence dans une tâche d'orientation vers une menace.

Les résultats obtenus sont en accord avec les données provenant d'expérimentations similaires rapportées par d'autres auteurs, qui concluent également en une amélioration de performances lorsque deux informations, visuelle et sonore, sont présentées simultanément.

Des recherches dans le domaine des caractéristiques de l'information sonore semblent indispensables pour améliorer l'intégration du son spatialisé dans les tâches de localisation d'objectif en aéronautique.

REMERCIEMENTS

Les auteurs remercient vivement les pilotes Français et Américains qui par leur coopération efficace ont rendu possible cette expérimentation, ainsi que MM. FALCO et CRIQUI (STTE) pour leur soutien, MM. PATIN, REAU et CAMINAL pour la mise en oeuvre du banc de simulation et l'organisation de l'expérimentation.

BIBLIOGRAPHIE

1- Begault D. R. (1993). Head-Up auditory displays for Traffic Collision Avoidance System advisories : A preliminary investigation. Human factors,35(4),707-717

2- Bronkhorst A.W., Veltman J.A., van Breda L. (1995). Application of a Three-Dimensional Auditory Display in a Flight Task. In press Human Factors

3- Leger A., Cursolle J.P., Leppert F. (1995). Viseur de casque et amélioration de la perception de la situation spatiale : approche expérimentale de l'intérêt et des limitations. AGARD-CP-575, AMP symposium on "Situation Awareness : Limitations and enhancement in the aviation environment", Brussels, belgium,24-27 April.

4- Pellieux L., Gulli C., Leroyer P., Piedecocq B., Léger A., Menu J.P. (1993). Approche expérimentale du son 3D : Méthodologie et résultats expérimentaux. AGARD-CP-541, AMP Symposium on "Virtual Interfaces: Research and Application", Lisbon, Portugal,18-22 Oct.

5- Posselt C., Schroter J., Opitz M., Divenyi P.L., Blauert J. (1986). Generation of binaural signals for research and home entertainment. 12ème congrès d'acoustique, Toronto.

6- Welch R..B, Warren D.H. (1986). Intersensory interactions. Handbook of Perception and Human Performance.

7- Wickens C., Vidulich M. (1982). S6R6C compatibility and dual task performance in two complex information processing tasks : Threat evaluation and fault diagnostic. Technical report EPL-82-3/ONR-82-3; University of Illinoisat Urbana Champaign.

8- Wightman F.L., Kistler D.J. (1989). Headphone simulation of free filed listening. Stimulus synthesis. Journal Acoust. Soc. Am. 85 (2) p858-867

Evaluation of a three-dimensional auditory display in simulated flight

A.W. Bronkhorst and J.A. Veltman
TNO Human Factors Research Institute, PO Box 23, 3769 ZG Soesterberg, The Netherlands

1. SUMMARY

Modern signal processing techniques allow headphone sounds to be processed in such a way that they seem to originate from virtual sound sources located in the three-dimensional space around the listener. By using head tracking devices, it is even possible to create a stable virtual acoustic environment that takes (head) movements of the listener into account. One interesting application of 3D sound is that it can be used to support situational awareness by generating virtual sound sources that indicate positions of relevant objects (e.g. targets or threats). This application was investigated in two flight simulation experiments in which the 3D auditory display, used as a radar display, was compared with 2D and 3D visual radar displays. A target localization task was employed, in which the subject, who flew a fighter aircraft, had to locate and follow another, suddenly appearing aircraft as quickly as possible. Dependent variables were the search time and a subjective workload score, obtained after each trial. In the second experiment, also the deviation from the optimal track toward the target and the performance on a secondary task were scored. Results show that search times and workload are similar for 3D auditory and 2D visual displays. Search times for the 3D visual display were smaller. Simultaneous presentation of auditory and visual displays gave clearly improved performance in case of the 2D visual display, but only minimal improvement with the 3D visual display. The results demonstrate the effectiveness of a 3D auditory display used as a radar display, but indicate that further development is required to reach the performance level of advanced 3D visual displays.

2. INTRODUCTION

The development of man machine interfaces for military aircraft is mainly focused on visual presentation of information and the use of new types of visual displays. Because relatively little attention is given to the auditory channel, the resulting auditory displays not only have poor ergonomics, in most cases, but also fall short of fully exploiting the information processing capabilities of the human auditory system. A major step forward, in this respect, is the recent development of techniques for three-dimensional (3D) sound presentation through headphones [1, 2, 3]. These techniques are based on simulation of the direction-dependent acoustic effects of the human body, head and ears through digital signal processing. By using 3D sound in an auditory display one can benefit not only from the human ability to localize sounds, but also from the internal noise suppression associated with binaural listening. The latter mechanism underlies the so-called cocktail-party effect: the ability to tune in on sounds coming from one direction while suppressing other sounds [4, 5].

Application of 3D sound within the cockpit of a military aircraft has three potential advantages. First, by presenting sounds from specific, meaningful directions (e.g. a threat warning sound from the direction where the threat has been detected) the situational awareness of the pilot can be supported. Second, communication efficiency can be improved by assigning different channels to (virtual) sources located at different points in space. Third, by presenting auditory signals from locations that are spatially separated, their detection and identification can be facilitated.

At the TNO Human Factors Research Institute, flight simulation experiments have been performed to quantify the advantages of 3D sound presentation with respect to situational awareness. The experiments used the context of a fighter jet cockpit and a task in which the pilot had to locate and trail a target aircraft, that appeared suddenly at an unknown position, as quickly as possible. The position of the target aircraft was indicated either auditorily, using 3D sound, or visually, on a radar display. By using a head tracker that covered all angular degrees of freedom, it was ensured that the 3D sound always pointed at the target, irrespective of the head orientation of the pilot. Two types of visual displays were evaluated: a conventional 2D outside-in display and an advanced 3D inside-out display. Task performance was quantified by determining the search time and the deviation from the optimal flight path. To enhance the workload, a secondary task was included during part of the conditions. The performance on this task was also evaluated.

3. METHODS

3.1 Experiments

Two flight simulation experiments were conducted. In experiment I, a 3D auditory and a 2D visual radar display were used. The latter display was modelled after the present threat warning display in fighter aircraft. Performance was determined without the radar displays, with either display presented alone, and with both displays presented simultaneously. In all conditions, subjects were presented with a high-resolution outside image, which showed the target

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

only at close range, and a tactical display, which indicated the position of the target at all ranges but only within a limited field of view. The 3D auditory display used in this experiment was based on measurements of acoustic transfer functions (head-related transfer functions or HRTFs) performed for each subject individually. The results of this experiments have been reported previously [6].

In experiment II, all three display types were employed: 3D audio, 3D visual and 2D visual. The outside image and tactical display were virtually the same as in the first experiment. However, a head-up display (HUD), indicating speed and heading, was added in order to stimulate the subjects to look at the outside image rather than the tactical display. As applications of 3D auditory displays will most probably use presentation in combination with a visual display, only conditions with visual and audiovisual radar displays were considered in this experiment. During half of the conditions, subjects had to perform a secondary task, which required them to press a button when a marker in the HUD crossed certain limits. In this experiment, the HRTFs employed in the 3D auditory display were not measured individually, but selected from 27 predefined sets, based on previous measurements for 10 subjects. Each subject for the flight simulation experiment was subjected to a listening test to determine for which set localization performance was optimal.

3.2 Subjects

Subjects were professional helicopter pilots and observers, also experienced in flying helicopters, employed by the Royal Netherlands Air Force. Their age ranged from 22 to 31. It was verified that all subjects had normal hearing at both ears: their hearing thresholds at octave frequencies from 250 to 8000 Hz were at most 20 dB (*re* ISO 386). Eight subjects participated in experiment I; 12 in experiment II.

3.3 3D sound generation

In order to generate spatialized sounds, the headphone signals were convolved in real time with digital filters. These filters were calculated in the frequency domain by multiplying the HRTFs, i.e. the transfer functions from an external sound source to eardrums, with the inverse transfer function from headphone to eardrum. The HRTFs were determined for approximately 1000 angles of incidence, covering almost the complete sphere around the listener with a resolution of $5-6°$. No interpolation between HRTFs was performed: when a certain direction was to be simulated, the HRTFs for the nearest measured position were used. The simulated direction was based on (1) the relative position of the target aircraft; (2) the orientation of the subject's aircraft and (3) the head orientation of the subject, as measured with a Polhemus Isotrak head tracker.

Measurement of the HRTFs were performed in the anechoic room of the TNO Human Factors Research Institute. The subjects were seated on a chair with a headrest and wore a headband equipped with a head tracker. Their head was located in the centre of an aluminium arc on which a trolley with a loudspeaker was mounted. A computer, controlling the positions of arc and trolley, ensured that the loudspeaker was placed subsequently at each of the 1000 predefined positions, taking into account the head orientation measured by the head tracker. At each position, a series of time-stretched pulses were emitted by the loudspeaker and recorded with probe microphones, placed in the ear canals of the subjects. The recordings were averaged, transferred to the frequency domain and then stored. A similar procedure was used for the determination of the headphone-to-eardrum transfer function. In this case, the pulses were generated by Sennheiser HD 530 headphones placed over the ears of the subject.

As indicated above, individual HRTF measurements were only performed for the subjects of experiment I. In these measurements, the tip of the probe tube was at a distance of several millimetres from the eardrum and the response at higher frequencies was somewhat influenced by the quarter-wavelength anti-resonance. Listening experiments showed that the virtual sources generated with these HRTFs could be located almost as accurately as real sources when head movements were allowed [7]. However, with the head fixed, significantly more localization errors were observed for virtual sources than for real sources.

The HRTFs used in experiment II were derived from a database of HRTFs for 10 subjects, measured much closer to the eardrum (within 1 mm). Subjective evaluation indicated that localization performance for these HRTFs approximated that for real sources (when the subjects used their own HRTFs). By applying first a Principal Component Analysis (PCA) to the HRTFs of each subject's ear and then a second PCA to the resulting data across ears and subjects, the dimensionality of the interindividual differences in the HRTFs was reduced to 3. Subsequently, a grid of $3 \times 3 \times 3$ points was constructed in this space, approximately covering the projected points of the 20 measured ears. For each ear of each subject, one of the 27 sets of HRTFs was selected. The selection was based on the results of a localization test, which was similar to the 'confusion' experiment described in Bronkhorst [7]. The results indicate that the localization performance achieved with these HRTFs is similar to that shown by the subjects of experiment I (but still poorer than the performance for real sources).

3.4 Flight simulation

The flight simulator consisted of a simple mock-up of a fighter cockpit with a computer screen containing all visual displays (except the HUD) and a 156° (horizontal) by 42° (vertical) outside image generated by an ESIG 2000 graphical processor. The controls available to the pilot were a throttle, a force stick for setting the pitch and roll, and a speed brake. The computer screen showed the altitude, speed, climb speed and compass heading of the aircraft. In addition, it contained a tactical display indicating the ground and the position of the target aircraft within a limited field of view (see Fig. 1). The 2D and 3D visual displays were also shown on the screen. In experiment I, indications of the heading and pitch were displayed as well; in experiment II these were shown on the HUD.

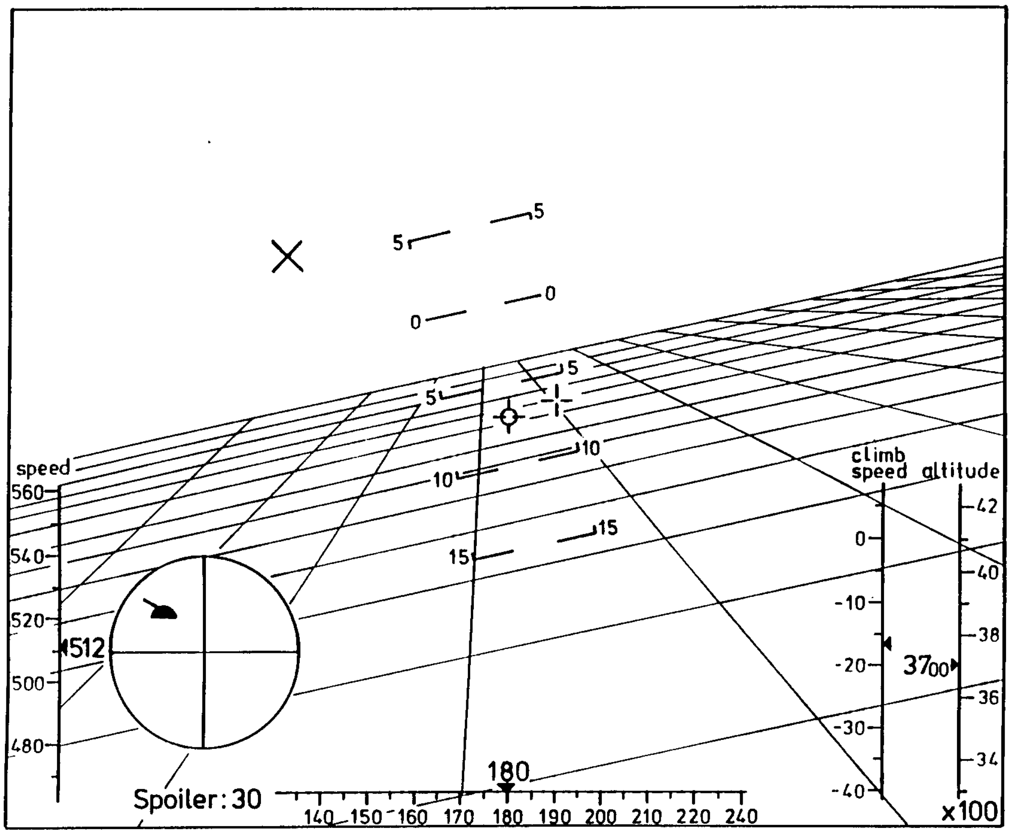


Fig. 1. The main display containing information on the status of the aircraft and a tactical display. The 2D radar display, used in experiment I, is shown in the bottom left-hand corner.

3.5 Radar displays

The 2D visual display consisted of a bird's-eye-view (outside-in) radar display, oriented heading up, containing a half-circular symbol and a line extending from the symbol. These indicated the relative position and speed of the target, respectively, as projected on the horizontal plane through the aircraft. The relative pitch of the target was indicated either by the colour of the symbol (in experiment I) or by a scale projected next to the display (in experiment II). The 3D visual display showed an inside-out image. The target position was indicated on a projected globe around the subject's aircraft (which was marked by a stylized symbol in the centre of the globe). Circles on this globe marked the horizontal plane and the plane through the wings of the aircraft. The target distance was shown on a separate scale. Both visual displays are illustrated in Fig. 2. The 3D auditory display generated a pulsed harmonic tone from the relative direction of the target. The harmonic tone contained components up to 15 kHz. Its level varied within a range of 10 dB (in experiment I) or 15 dB (in experiment II), depending on the relative distance of the target.

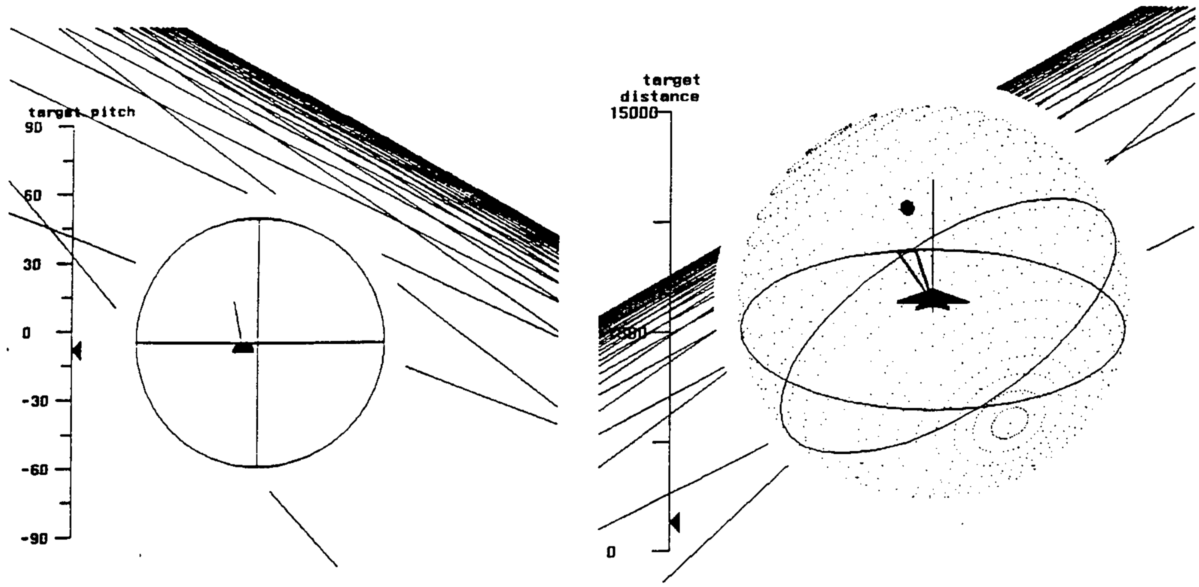


Fig. 2. The 2D and 3D visual radar displays used in experiment II.

3.6 Conditions and tasks

There were four conditions in experiment I: without radar display, with the 2D visual display, with the 3D auditory display and with both displays. In experiment II, one of the visual radar displays (2D or 3D) was always presented and the 3D auditory display was either added or left away. In addition, all display configurations were tested with and without secondary task. This resulted in a total of eight experimental conditions. Each condition consisted of 18 to 20 trials. At the start of each trial, the subject followed the target aircraft, which flew a fixed route with constant speed. At a certain moment, the target disappeared and reemerged at an unknown position. The task of the subject was to locate and follow the target, and to keep it in front of the own aircraft within an angle of $\pm 10°$. In experiment I, the requirement was that this limit should not be exceeded during 5 s. As this often resulted in rather long search times, it was decided to skip this requirement in experiment II and to end the trial either when the target was within $\pm 10°$ or when the search took more than 20 s. After each trial (in experiment I) or condition (in experiment II), the subject was asked to indicate the subjective workload that he or she had experienced on a rating scale. This scale ranges from 0 to 150 and provides nine labelled anchor points. The secondary task required the subject to press a button whenever a marker, sliding randomly along a scale, crossed an upper or lower limit. The marker also changed colours when either limit was exceeded. Marker and scale were projected on the HUD. All subjects participated in training sessions for approximately five hours before the actual experiment was conducted.

The independent variables in experiment I were the search time and the workload score. In experiment II, several additional variables were determined: the deviation from the optimal flight path (the tracking error), the percentage of targets not found within the maximum duration of the trial, as well as the response time and miss rate for the secondary task. The tracking error was defined as the angle between the plane through the longitudinal axis of the aircraft that coincides with the flight path and the plane through the same axis that contains the target.

4. RESULTS

The average search times and workload scores obtained in experiment I are shown in Fig. 3. It appears that the reduction in search time with respect to the no-display condition is approximately the same for both radar displays.

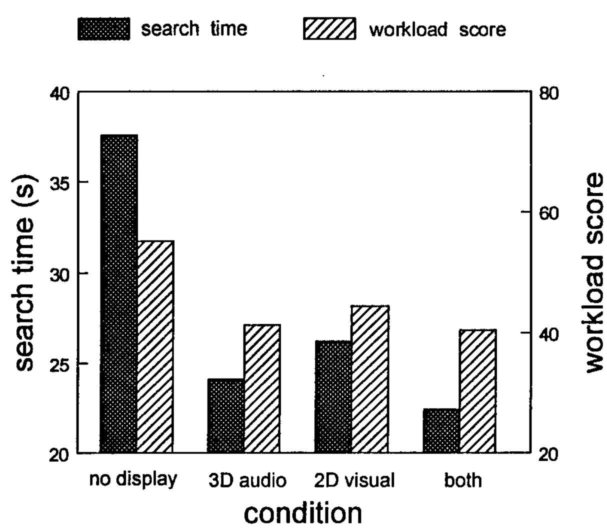


Fig. 3. Average search times and workload scores obtained in experiment I.

A further, significant, reduction occurs when the two displays are presented simultaneously ($p<0.01$). The workload scores indicate that subjects had relatively little difficulty using the radar displays. Furthermore, it appears that the improvement in performance for the condition with both displays does not occur at the cost of a higher workload.

Search times were shorter in experiment II than in experiment I because there was no requirement with respect to the time that the subject should stay behind the target. In Fig. 4, the average search times and tracking errors for the eight conditions are displayed. The results demonstrate a main effect of visual display type: both search time and tracking error are considerably shorter for the 3D than for the 2D radar display. When the 3D auditory display is presented as well, a small but significant reduction of search time and tracking error occurs for the 2D visual display ($p<0.05$). The percentage of missed targets (not shown in the figure), is affected even more: it drops from 32% to 23%. For the 3D visual display, only the tracking error is reduced significantly when the 3D auditory display is added ($p<0.05$). Neither search time nor the percentage of missed targets are affected. (The observed increase in search time for the combined 3D auditory and visual displays is not statistically significant.) It further appears that the presence of the secondary task has a significant effect on the search time ($p<0.01$) but not on the tracking error. This effect does not interact with the visual display type nor with the presence of the auditory display.

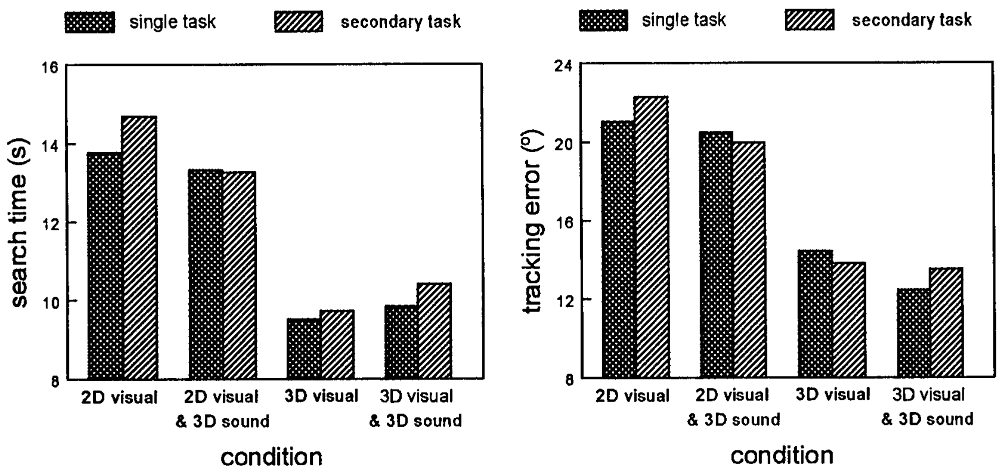


Fig. 4. Average search times and tracking errors for the eight conditions of experiment II.

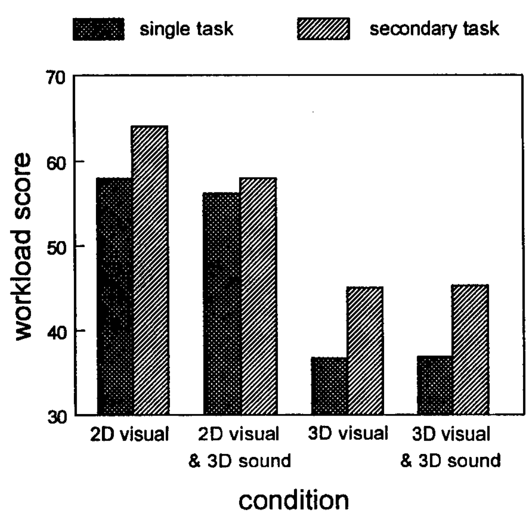


Fig. 5. Average workload scores obtained in experiment II.

Analysis of the results for the secondary task reveal significant effects of visual display type and the presence of the 3D auditory display on the miss rate: less events are missed with the 3D visual display than with the 2D display ($p<0.01$), and the number of misses decreases further when the 3D auditory display is added ($p<0.05$). Reaction times were, however, not affected significantly. The subjective workload scores for the eight conditions are shown in Fig. 5. There are main effects of visual display type ($p<0.001$) and secondary task ($p<0.01$); addition of the 3D auditory display did not affect the workload.

5. DISCUSSION

The flight simulation experiments show that a 3D auditory display, that indicates the position of a target to be located by generating a warning tone from its relative direction, is equally effective as a conventional visual display, showing an outside-in 2D radar image as well as the relative pitch of the target. Furthermore, the performance for the visual display improves when the 3D auditory display is presented simultaneously. Though these results demonstrate the potential value of a 3D auditory display, used as radar display, it appears from the comparison with the 3D visual display that there is still considerable room for improvement. Performance with the latter display is clearly better than with the combined 2D visual−3D auditory display. Interestingly, there is still an effect of adding the 3D auditory to the 3D visual display, but the resulting performance improvement is only small.

The results, thus, show that the 3D auditory display used in the present experiments has only limited effectiveness, compared with an advanced 3D visual display. It appeared from an evaluation of reactions given by subjects that the directions indicated by the auditory display were often not recognized immediately, or they were misinterpreted due to front-back or up-down confusions. As research has shown that it is, in principle, possible to achieve accurate sound localization with only a minimum number of confusions, even when no head movements are possible [2, 8], it must be concluded that performance of the subjects was affected by the limitations of the present 3D auditory display. This demonstrates that for demanding applications, it is required to use a high-quality 3D auditory display, optimally adapted to the individual user.

A second aspect that should be considered when comparing 3D auditory and visual displays is that the auditory display can be further improved by using not only 3D sound, but also specific, meaningful signals, possibly adapted to the task for which the display should be used. Such an optimization of the symbology was, in fact, already performed for the present 3D visual display, which is based on a large body of literature and which has been evaluated in previous flight simulation experiments. In the auditory display, one could, for example, use (slightly) different signals for different hemispheres or quadrants, in order to prevent front-back or other confusions. Alternatively, an important parameter for the task, like the tracking error in the present search task, could be coded into the signal. Thus, in developing 3D auditory display for cockpit applications, one should not only aim at a correct simulation of the directional hearing, but one should also pay attention to a careful design of the auditory symbology.

6. REFERENCES

1. Wenzel, E.M., Wightman, F.L., and Foster, S.H., "A virtual display system for conveying three-dimensional acoustic information", *in* "Proceedings of the Human Factors Society 32nd Annual Meeting", Santa Monica, CA, Human Factors and Ergonomics Society, 1988, pp. 86-90.
2. Wightman, F.L., and Kistler, D.J., "Headphone simulation of free-field listening: II. Psychophysical validation", J. Acoust. Soc. Am., 85, 1989, pp. 868-878.
3. Begault, D.R., "Head-up auditory displays for traffic collision avoidance system advisories: A preliminary investigation", Human Factors, 35, 1993, pp. 707-717.
4. Cherry, E.C., "Some experiments on the recognition of speech, with one and with two ears", J. Acoust. Soc. Am., 25, 1953, pp. 975-979.
5. Bronkhorst, A.W., and Plomp, R., "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing", J. Acoust. Soc. Am., 83, 1992, pp. 1508-1516.
6. Bronkhorst, A.W., Veltman, J.A., and Van Breda, L., "Application of a three-dimensional auditory display in a flight task", Human Factors, 38, 1996, pp. 23-33.
7. Bronkhorst, A.W., "Localization of real and virtual sound sources", J. Acoust. Soc. Am., 98, 1995, pp. 2542-2553.
8. Oldfield, S.R., and Parker, S.P.A., "Acuity of sound localization: a topography of auditory space. I. Normal hearing conditions", Perception, 13, 1984, pp. 581-600.

Flight Demonstration of an Integrated 3-D Auditory Display for Communications, Threat Warning, and Targeting

Richard L. McKinley, Capt. William R. D'Angelo, Mark A. Ericson
Crew Systems Directorate, Armstrong Laboratory
2610 Seventh Street, Building 441
Wright-Patterson Air Force Base, OH 45433-7901 USA

1. SUMMARY

Recent laboratory experiments have demonstrated significant increases in visual target acquisition performance when the subjects have been aided by 3-D audio cueing. The USAF Armstrong Laboratory's 3-D audio display system was integrated with a helmet mounted display system on a Navy/Marine TAV-8B Harrier for a joint Air Force/Navy flight demonstration. The 3-D audio system has the capability of synthesizing signals that when presented over headphones give the user the illusion that the sound is emanating from some external location. These synthesized signals can be configured to emanate from selected known sources to indicate their location on the ground and in the air. Military aircraft applications of 3-D audio include threat location warning, wingman location indication, spatially separated multi-channel communications, and audio target location indications. For this flight demonstration, the Armstrong Laboratory 3-D audio system implemented spatially separated communications, threat location cueing, and target location aiding. Laboratory experiments of combined audio-visual search performance resulted in target acquisition time reductions of approximate 50 percent and workload reductions of approximately 20 percent. In March 1996, the data collection portion of the flight demonstration was initiated. The integration of the 3-D audio system into the TAV-8B, the laboratory experimental results, and the preliminary results of the flight demonstration are presented in addition to recommendations for future research and flight tests.

2. Introduction

The auditory modality is the only true full coverage 3-D human sensory system. All other human sensory systems rely on multiple samples of portions of the surrounding space. Humans, with fully functional auditory systems depend on that system to provide warning and cueing for events outside their current field of view or interest. If a significant auditory event occurs, the first and immediate reaction is to turn the eyes, head, torso, and/or body to bring the high resolution, or foveal, portion of vision to bear on the spatial area of interest. In essence, the auditory modality acts as an early warning system for the organism. The auditory system also provides high speed control input to the high resolution but slower visual system for better "threat or target assessment." This interaction between the auditory and visual system is used by almost all humans every day. Current aircraft cockpit audio systems are not capable of presenting auditory signals which have this important spatial information. Recently, audio signal processing technology has been developed to allow the presentation of spatial audio information over headphones. This technology allows pilots in the cockpit to

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

utilize this synergistic interaction of auditory and visual sensory modalities for a wide range of applications including threat and target spatial localization and assessment.

3. Background

The United States Air Force's Armstrong Laboratory began the development of a 3-D audio display system in 1985. The first working model of a real-time, head-motion coupled, 3-D audio localization cue synthesizer was demonstrated by McKinley in February, 1987 (1). This system produced auditory signals that when presented to a human listener via headphones, were perceived to be originating at a given location external to the listener. This auditory perception was immediate, intuitive, and required no training. The results of an initial series of experiments (2), demonstrated that synthetic auditory cues could be localized with an average error of 4-8 degrees as shown in Figures 1-2, both in quiet and in cockpit noise laboratory environments.

The 1992 joint Air Force and Navy 3-D audio flight demonstration, sponsored by the Advanced Research Projects Agency, demonstrated in-flight functionality of the 3-D audio concept and accomplished the first integration of the 3-D audio system with an operational flight platform (3, 4). The flight demonstration used the 3-D audio system to indicate to the pilots the location of the preplanned ground targets and to spatially separate the two radio channels available on the AV-8B. Overall, the results of the demonstration were successful and led to additional laboratory studies investigating the effects that 3-D audio has on target acquisition and detection.

These additional laboratory and simulator studies demonstrated significant improvements in user performance in speech communications, target acquisition, and target detection. Improved speech intelligibility was demonstrated by virtually separating talkers using the 3-D audio system (5), sample data is shown in Figure 3. The simultaneous voice communications of two radio channels were spatially separated. Channel 1 was located 45 degrees to the left and channel 2 was located 45 degrees to the right of the listener, both at 0 degrees elevation. An average increase in the intelligibility of the voice communications of more than 25 percent was measured in a noise field of 105 decibels (dB), Sound Pressure Level (SPL). A simulator study demonstrated that visual target acquisition times were reduced by up to 50 percent, as shown in Figures 4-5, and target detection times reduced by 50 to 100 percent, as shown in Figures 6-7, when the 3-D audio displays were integrated with the visual search procedure to create an audio-visual search method (6, 7). These studies demonstrated the potential for enhanced aviator performance.

4. Objective/Approach

The objective of this paper is to describe the application and integration of 3-D auditory display technology for communications, threat warning, and targeting in a high performance tactical fighter aircraft environment. The approach was to integrate the Armstrong Laboratory 3-D auditory system and other pilot-vehicle-interface (PVI) systems with an AV-8B Harrier aircraft and conduct a series of flight demonstrations to subjectively measure the benefits of an integrated system.

5. Equipment

For the flight demonstration, the 3-D audio system, shown in Figure 8, was mechanized to provide cues for several different cockpit functions:

Enhanced Voice Communications. This application of 3-D audio was implemented to improve voice communications intelligibility and situational awareness (SA) in high workload that resulted in high stress conditions. The voice communications of the two radios were spatially separated 45 degrees left and 45 degrees right of the pilot to achieve the advantages observed in the earlier Harrier flight study.

Navigation Aid. The customized audio symbol to aid pilot navigation was a swept sine wave, amplitude envelope modulated to sound like a water drop. This waveform was used as the basis for three different auditory symbols. The first symbol, a single water drop, was used to cue the location of a designated waypoint. The second symbol, 5-second train of water drops, was used to indicate the direction to the next waypoint; it was activated using the COURSE voice command of the Interactive Voice Module (IVM). The third symbol, continuous water drops, was used to aid orientation whenever the Targeting POD (TPOD) video was displayed on the Helmet Mounted Display (HMD). The audio cue indicated where the TPOD was looking, a direction that was most likely radically different from both where the pilot was looking and the heading of the aircraft.

Radar Warning Receiver (RWR). The 3-D audio system was configured to add location information to threat warning tones of the RWR. This spatial information was integrated with the warning tones of the top four highest priority threats. The audio symbology was designed and developed jointly by a team consisting of the Georgia Tech Research Institute (GTRI), the Armstrong Laboratory, and the flight test pilots. Pilot training time was minimized by starting with the AV-8B RWR tones and implementing a priority scheme based on volume and the perceived pitch of the audio cues. Additionally, the spectral content of the standard tone set was broadened by the addition of a low level broad spectrum signal to increase localization accuracy.

The 3-D audio system hardware consisted of a 3-D audio display generator, control panel, and Bose, stereo Active Noise Reduction (ANR) earcups. The generator weighed ten pounds, was $6.9 \times 5 \times 9.75$ inches, and was mounted in the aft cockpit. The generator communicated with the aircraft Mission Computer (MC) over the 1553 multiplex bus and with the Navy Standard Magnetic Tracker (NSMT) over an RS-422 line. Audio signals were received from communication (COMM) channels 1 and 2 and from the Auxiliary Communication Navigation and Identification Panel (ACNIP). The control panel weighed 2.3 pounds, was $3.75 \times 5.75 \times 5$ inches and was mounted in the fore cockpit, in front and to the right of the pilot. The control panel contained four independent volume controls and a master volume control. Modes of operation of the 3-D audio display were also switched using the control panel.

ANR technology reduced the overall level of noise that reached the ear of the earcup wearer by employing the phenomenon of destructive interference of sound waves. A miniature microphone located inside the earcup measured the noise field at the ear. The system electronics processed and inverted the noise signal phase and returned the noise

signal to the inside of the earcup at approximately the same level as the original noise. The acoustic delay of the system limited the maximum frequency of the active noise cancellation to about 1500 Hertz (Hz) and below. The passive attenuation of the earcup-helmet system was poor at low frequencies but increased to a maximum of about 38 dB at 8000 Hz. The ANR active attenuation was directly added to the passive attenuation providing about 7 dB more attenuation at 31 Hz and a maximum of about 21 dB more attenuation at 200 Hz.

The militarized ANR earcup was designed to replace the standard earcup of the HGU-55/P and HGU-56/P flight helmets. Installation into the flight test helmet (a modified HGU-55/P) involved exchanging earcups and wiring harness. The ANR earcup featured an original design silicon gel cushion that provided an acoustic seal. Each ANR earcup also contained significant shielding and filtering components that minimized the radiation and conduction of electromagnetic interference (EMI). Each earcup weighed 0.45 pounds and the system was wired for stereo inputs.

6. Method

Flight demonstrations were conducted using a two seat model of the AV-8B Harrier vertical take-off and landing attack aircraft. Subjective data were collected on situational awareness and workload during the individual trials and immediately following the completion of a trial. A trial was a single flight pass simulating the attack of a target. Several trials were completed during a 60-90 minute flight. The subject test pilot was seated in the forward cockpit of the AV-8B with a safety pilot in the aft cockpit. Audio and video recordings from the front cockpit were made of each flight.

7. Subjects

All subjects for the laboratory studies were paid volunteers and all had normal hearing thresholds (less than 15 dB Hearing Threshold Level, HTL) at each of the standard audiometric test frequencies 500 Hz, 1 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz. They were paid minimum wage plus a bonus for completing the study. The test subjects used in these studies were highly trained, performing approximately 4 hours per day, 5 days per week in a wide range of psychoacoustic and voice communication experiments. In all laboratory experiments the number of male and female subjects were equal.

Test subjects for the flight demonstration were Air Force, Navy, and Marine test pilots, one from each service. The hearing sensitivity for these three test pilots was no worse than 30 dB HTL at any of the audiometric frequencies.

8. Preliminary Results

The preliminary results of the flight demonstration showed that 3-D audio cueing, when integrated with aircraft sensor systems and other PVI technologies enhance pilot situational awareness and improved overall performance while reducing workload. The 3-D audio cueing for RWR allowed multiple threats to be managed while other flying tasks were being performed. The navigation waypoint cueing allow pilots to turn and be on course before having to look inside the cockpit at a navigation display. The auditory cueing of the TPOD look angle did not seem to contribute to pilot awareness and will need

additional research and development if this application is to be viable. The combination of ANR with the 3-D audio system was very well received. The ANR provided a significantly quieter work environment for the pilots and was judged to contribute to reduced workload and improved performance. The auditory symbology selected by the pilots was acceptable for tracking two simultaneous targets of the same type but was unsatisfactory for tracking three or four simultaneous targets of the same type.

12. Summary

This flight demonstration program showed the potential for 3-D audio technology to improve situational awareness, enhance performance, and reduce workload in a military high performance fighter aircraft environment. The specific integration of 3-D audio displays with aircraft sensor systems and other PVI technologies is critical to the overall performance of the system. Significant laboratory, simulator, and flight test research and development work remains in order to efficiently and effectively optimize the use of 3-D audio displays and other advanced audio technologies in aircraft of the future. The near term current focus of the research should include efforts on developing spatial auditory symbology. This new technology, 3-D audio displays, is very promising, it is up to the researchers to develop and produce the viable applications.

References

1. McKinley, Richard L., "Concept and Design of an Auditory Localization Cue Synthesizer," Air Force Institute of Technology, Masters Thesis, AFIT/GE/ENG/88D-29, 1988.

2. McKinley, Richard L., Ericson, Mark A., D'Angelo, William R., "3-Dimensional Auditory Displays: Development, Applications, and Performance," Aviation, Space, and Environmental Medicine, May, 1994.

3. Naval Air Warfare Center Weapons Division. "Airborne Demonstration of a 3-D Audio Display," by J. J. Martin. China Lake, CA, NAWCWPNS, June 1993. (NAWCWPNS TP 8138, publication UNCLASSIFIED).

4. Naval Air Warfare Center Weapons Division. "TAV-8B Demonstration Flight Debrief Summaries for Airborne 3-D Audio Cues," by M. P. Kibbe and D. J. Francis. China Lake CA, NAWCWPNS, September 1993. (NAWCWPNS TM 7586, publication UNCLASSIFIED)

5. Ericson, Mark A., and McKinley, Richard L., "The Intelligibility of Multiple Talkers Separated Spatially in Noise," *in* Binaural and Spatial Hearing in Real and Virtual Environments, eds. Gilkey and Anderson, Lawrence Erbaum and Associates, Mahwah, New Jersey, 1996.

6. Perrott, David R., Cisneros, John, McKinley, Richard L., and D'Angelo, William R., "Aurally Aided Visual Search Under Virtual and Free-Field Listening Conditions," Human Factors journal, in press.

7. McKinley, Richard L., D'Angelo, William R., Perrott, David, R., Nelson, W. Todd, Hettinger, Larry J., and Brickman, B. J., "An Initial Study of the Effects of 3-Dimensional Auditory Cueing on Visual Target Detection," Proceeding of the Human Factors and Ergonomics Society 39th Annual Meeting, pp. 119-123, 1995.

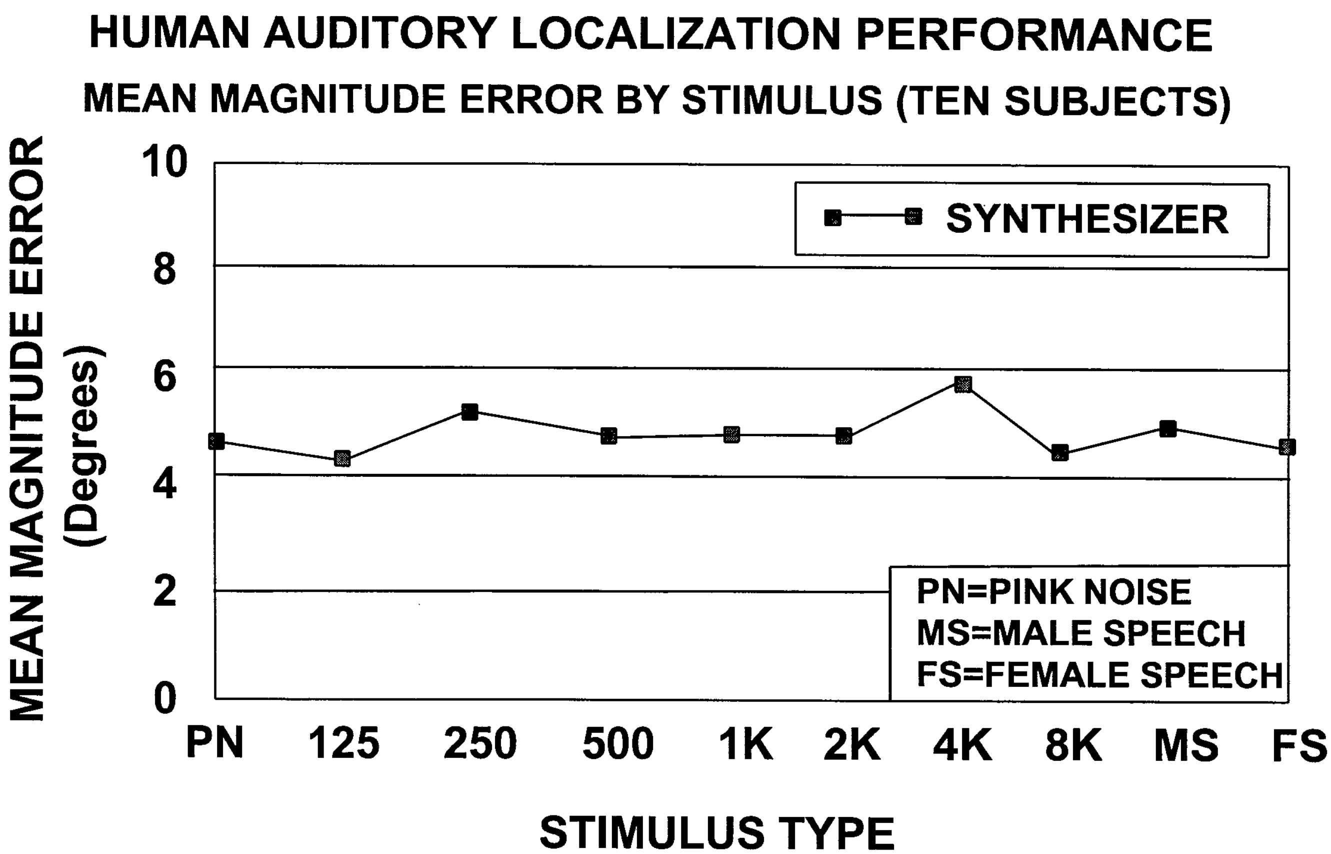


Figure 1. Human auditory localization performance in mean magnitude error by stimulus type for 10 subjects in quiet.

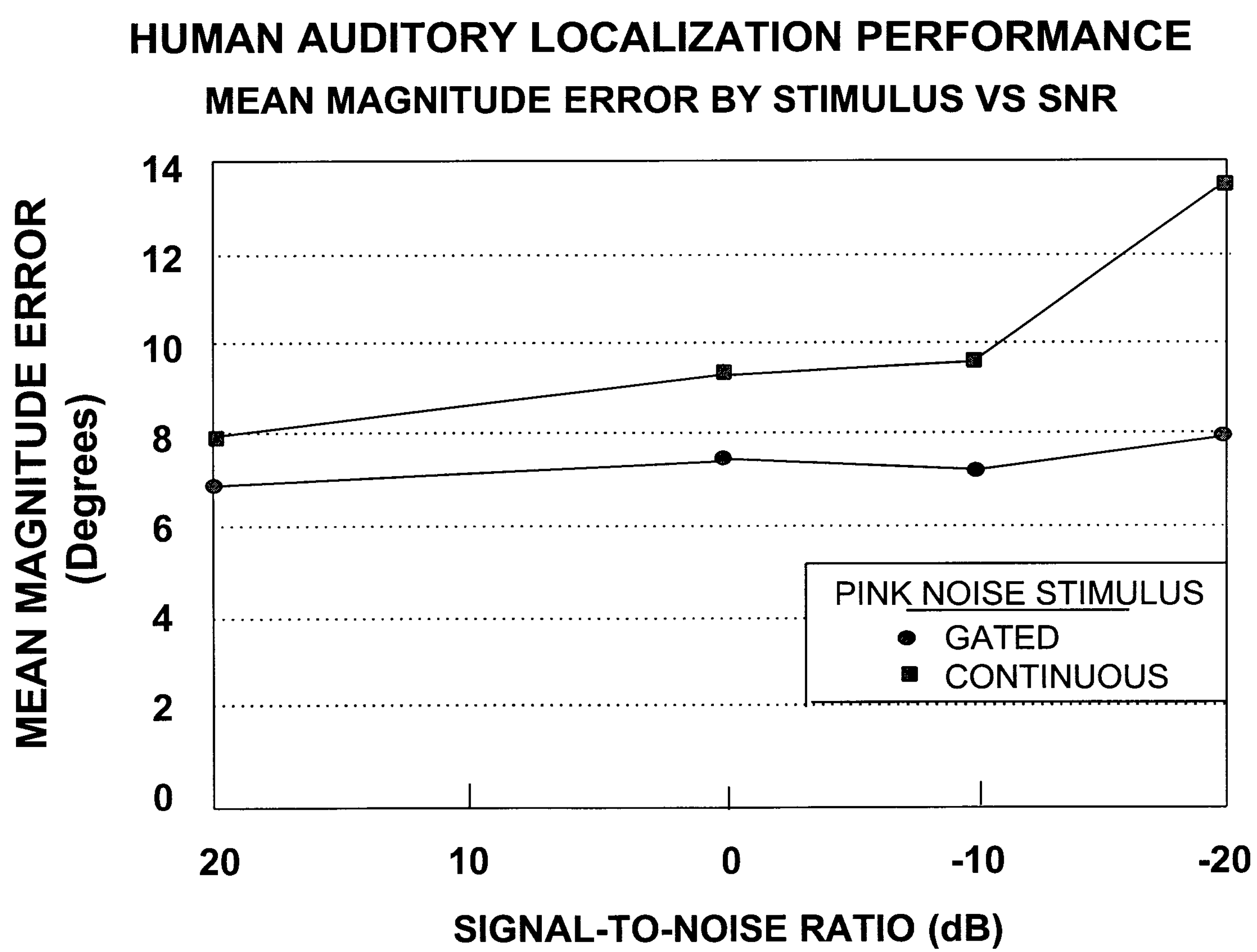


Figure 2. Human auditory localization performance in mean magnitude error with pulsed pink noise stimuli for 10 subjects in high level ambient pink noise.

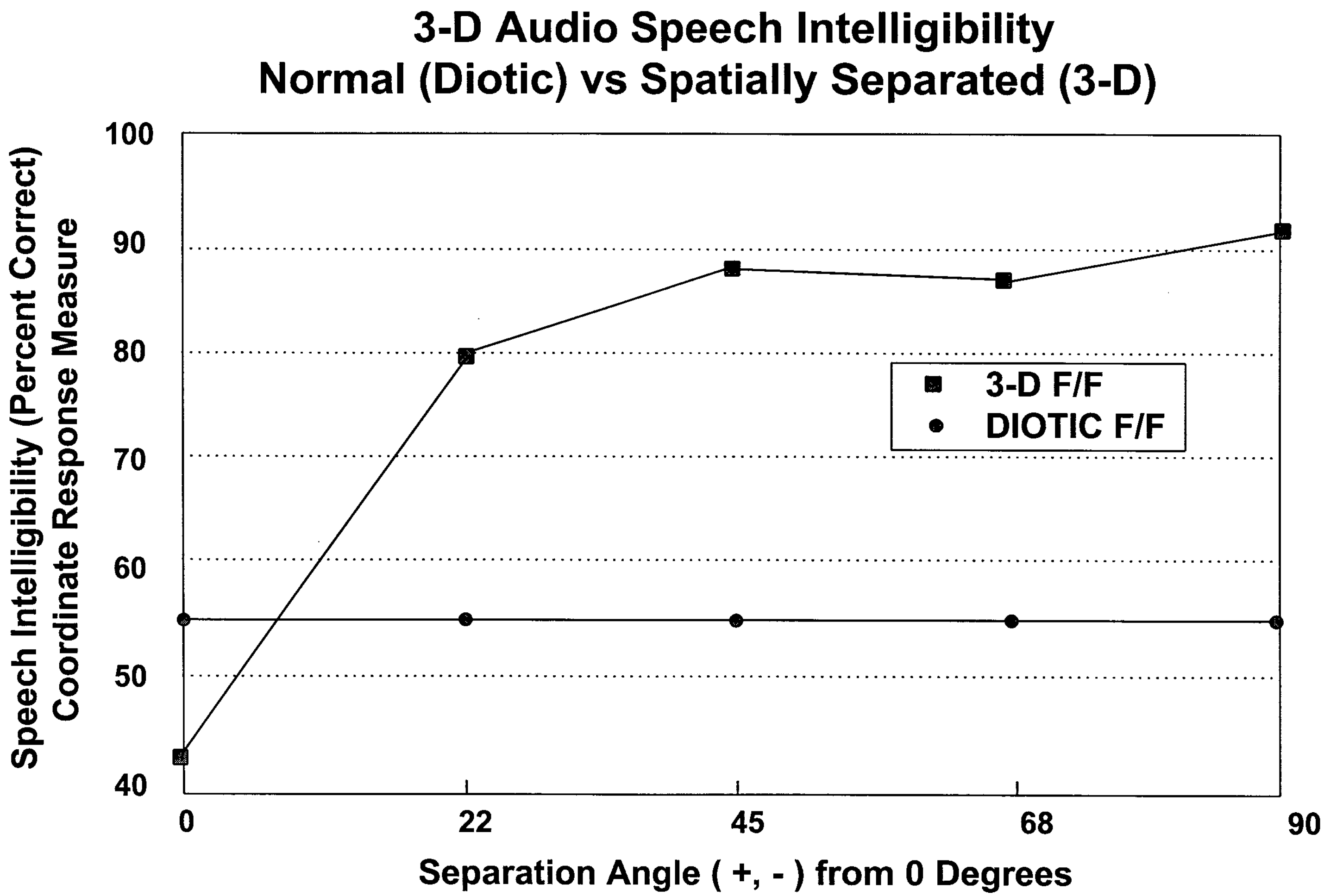


Figure 3. Average (3 pairs of 2 talkers x 4 listeners) coordinate response measure intelligibility for two competing messages in 115 dB ambient SPL pink noise for directional (spatially separated) with head motion vs diotic speech presentation at multiple angles of separation in azimuth.

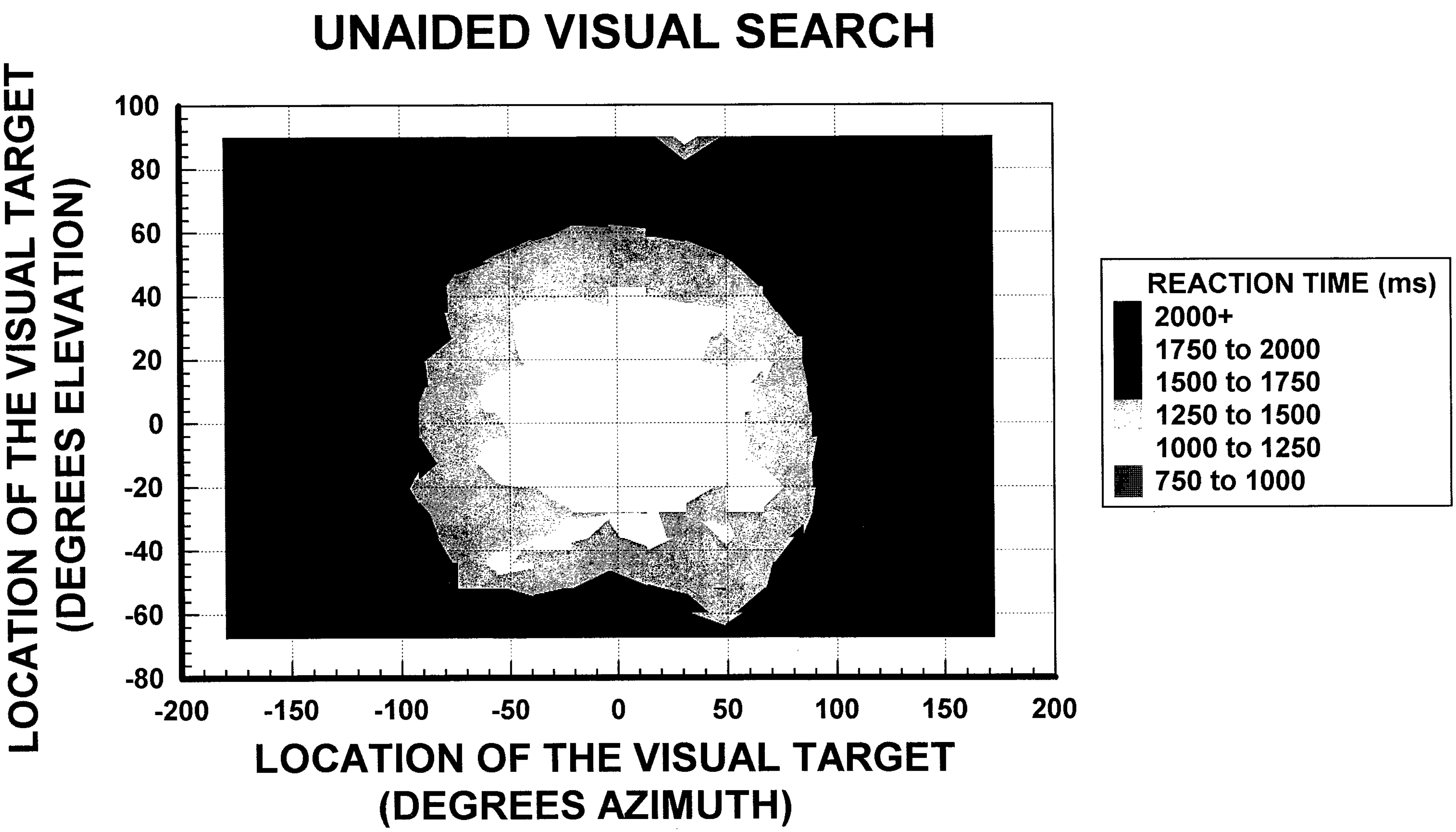


Figure 4. Visual target acquisition times (reaction times) using unaided visual search in azimuth and elevation.

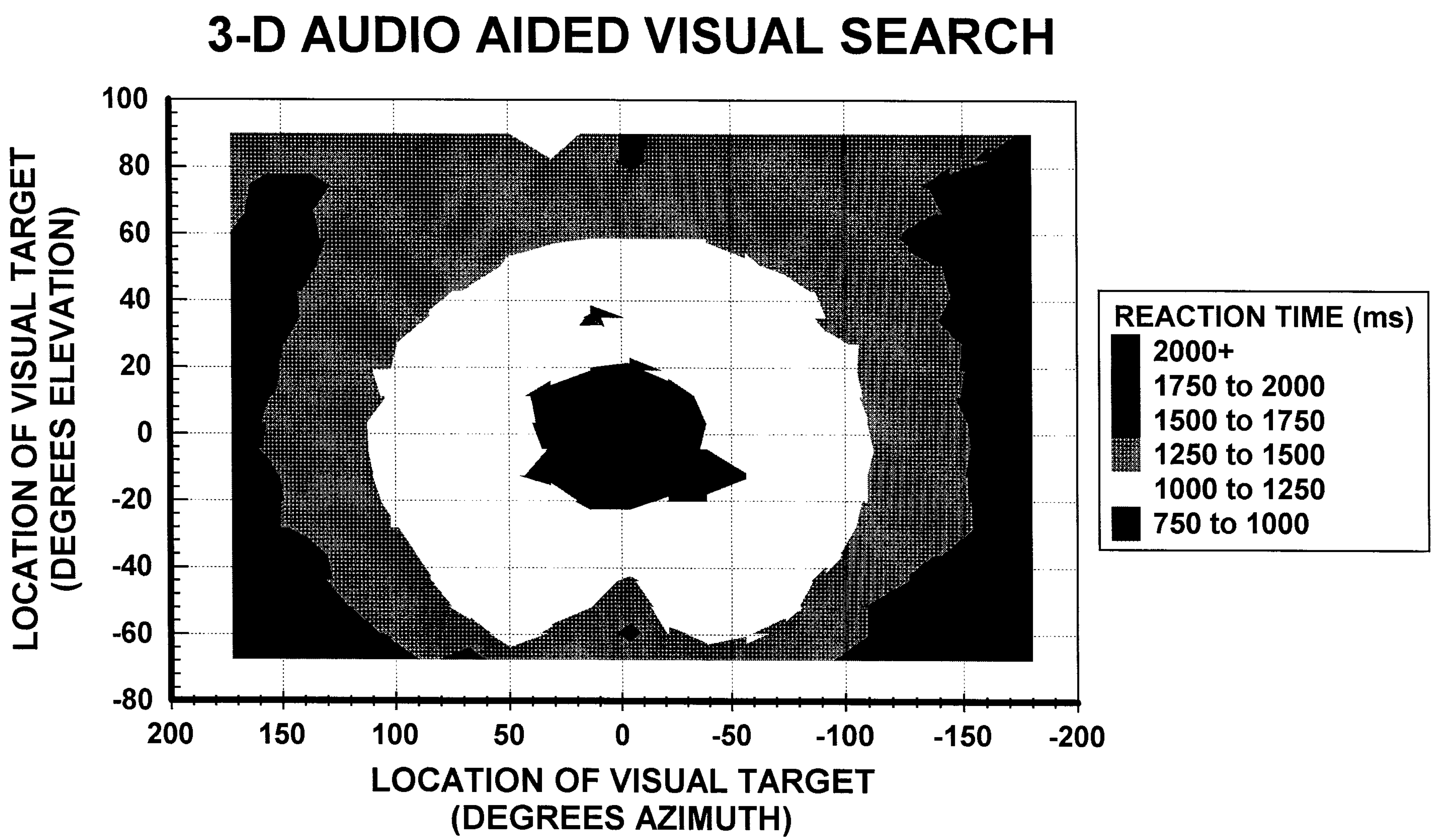


Figure 5. Visual search acquisition times (reaction times) aided by 3-dimensional audio displays.

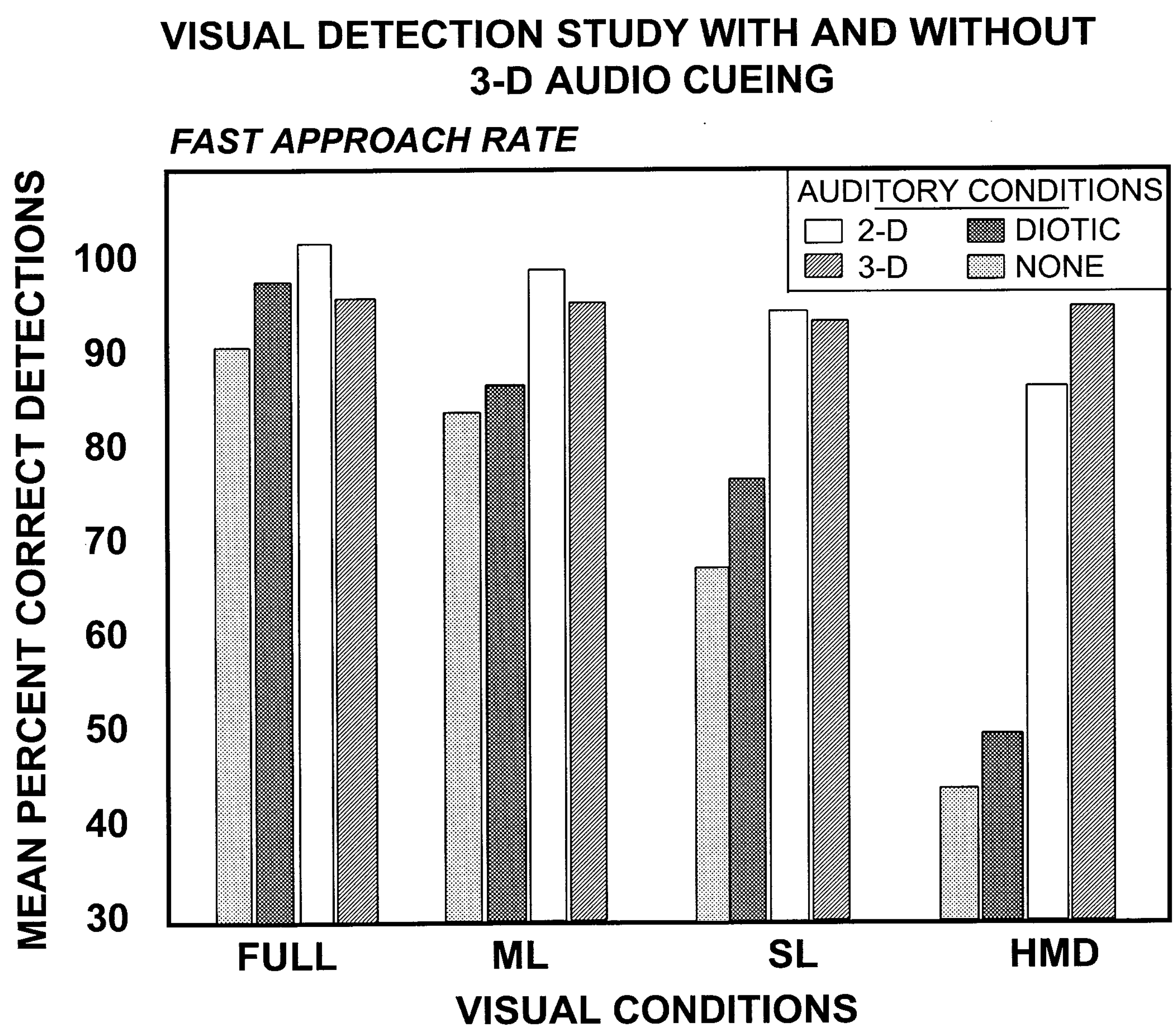


Figure 6. Mean percent correct visual detections of targets with and without 3-D audio cueing.

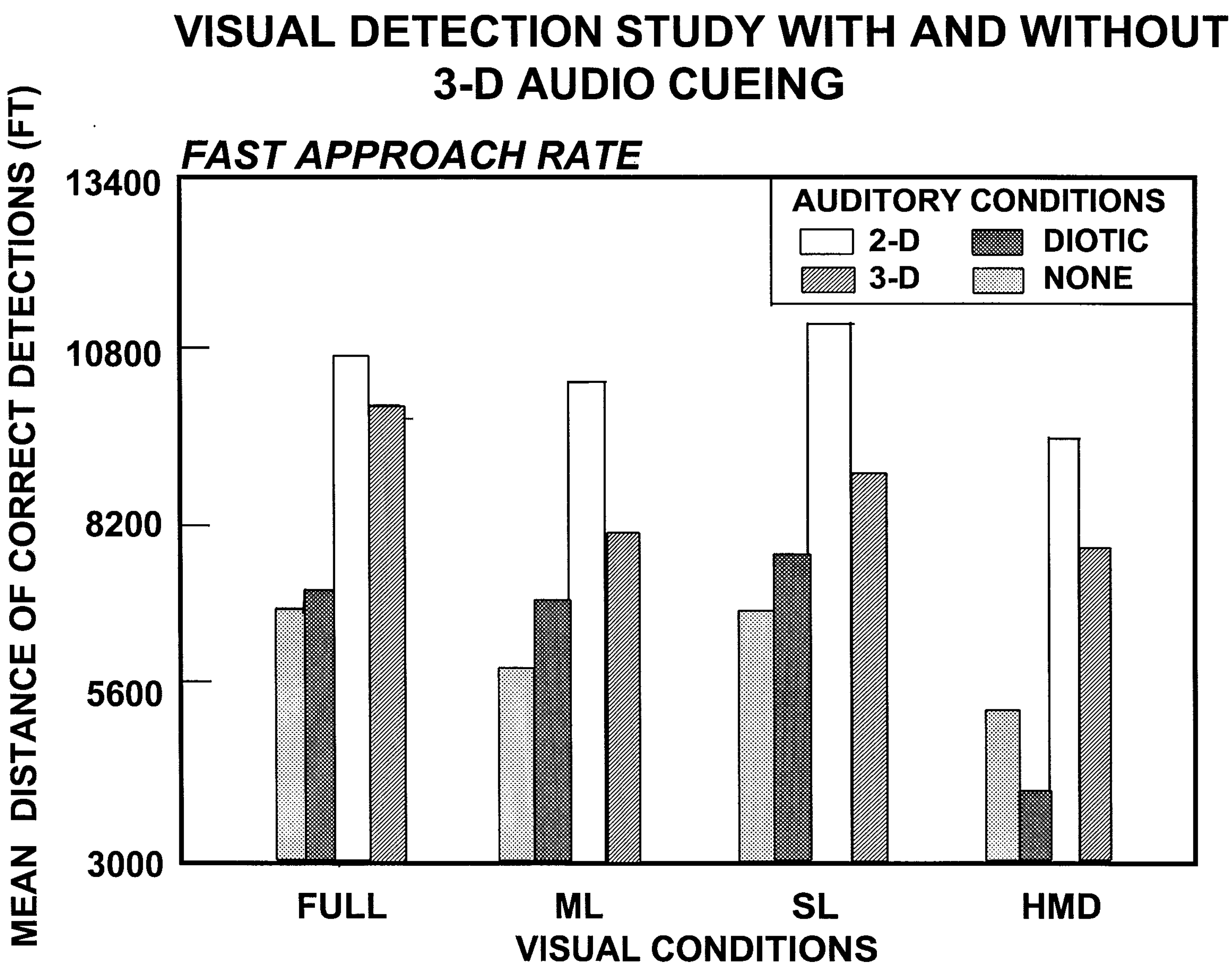


Figure 7. Mean distance of correct visual detections of targets with and without 3-D audio cueing.

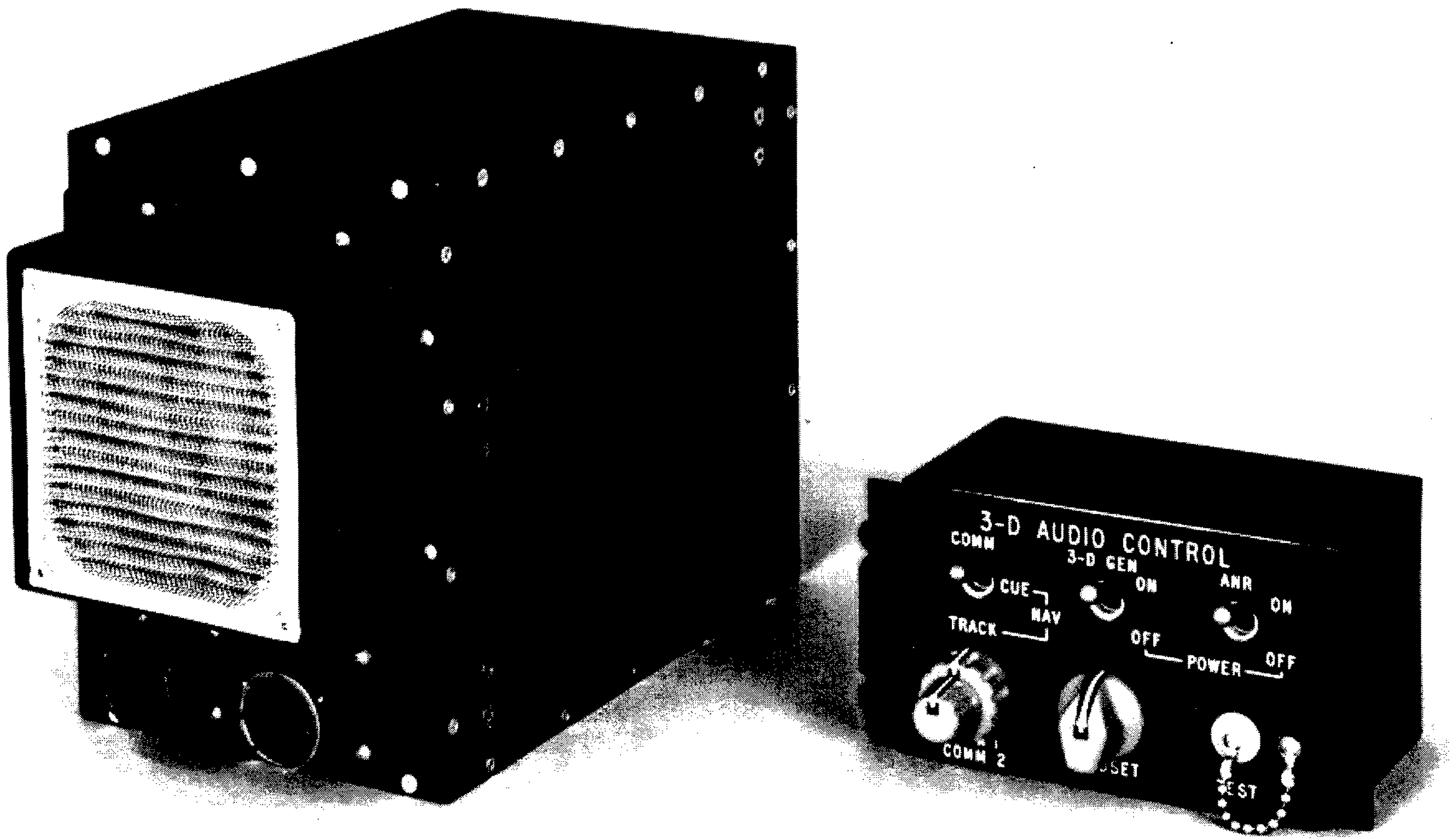


Figure 8. Flight worthy 3-D audio display system utilized during the flight demonstration. The processing system (left) was mounted in the fuselage and the control module (right) was mounted in the rear cockpit of the flight demonstration aircraft.

AUDIO WARNINGS FOR MILITARY AIRCRAFT

S.H.James

Systems Integration Department, Air Systems Sector
Defence Research Agency, Farnborough
Hampshire GU14 6TD, UK

SUMMARY

A survey of warning systems currently installed in military aircraft showed that generally they are of poor design. Simply constructed warning sounds have been added to aircraft as and when deemed necessary and hence, have been installed on an individual basis rather than as an integrated warning set. As more and more of these types of sounds are introduced discrimination will become more difficult and confusions increase. Additionally, the warnings are continuous in nature and presented at too high a volume which not only causes startle but interferes with communications, resulting in aircrew seeking the audio mute rather than dealing with the problem at hand. Consequently, the audio warnings currently in use may prove counter-productive and have flight safety implications.

This paper details the research conducted by DRA aimed at providing the UK military aircraft fleet with a standardised, fully integrated audio warning suite. To date the work has culminated in the development of a set of design guidelines and a presentation strategy that not only minimises the number of warning sounds required in a warning set but that remains flexible to allow new warnings to be added without necessarily increasing the number of sounds required. The characteristics for trend indicating sounds are also defined and a protocol for their design detailed. Additionally, in an attempt to enhance the number of audio alerts aircrew can process, manage and respond to accurately, the feasibility of mapping aircraft threat related warnings in three dimensional space is discussed and future research areas detailed.

1. INTRODUCTION

The warning systems that have traditionally alerted aircrew to problems that have arisen have mostly relied on visual signals in the form of warning lights on a central warning panel (CWP). However, during the early 1980s aircrew began flying regularly with Night Vision Goggles (NVGs) where the eyes are focused on the horizon and out of the cockpit for longer periods. This, coupled with their increasing operational workload enhanced the possibility that an illumination on the CWP may pass unnoticed. Consequently some aircraft began to introduce audio warnings as backup to the visual warning system.

A survey performed by the Defence Research Agency (DRA) in 1993[1] of the audio warnings currently used in the UK military fleet showed that in helicopters eleven different warnings are now backed by audio. Unfortunately, these warning sounds have been added to aircraft as and when a particular aircraft system, flight mode or threat has been deemed to require audio backup. Hence, they have been designed and installed on an individual basis rather than as an integrated warning set. The survey showed that there had been little or no apparent reference to other sounds already existing within or between aircraft types and highlighted a number of concerns arising from this ad hoc approach to audio warning implementation. Namely, the number and types of sounds currently installed and their audio presentation levels.

Currently no standardised audio warning implementation strategy exists and therefore the number of warning sounds is not limited, a new sound is simply added to the set when a new warning requires audio backup. The warning sounds used are of simple construction, either being alternating tones, frequency sweeps or repetitive bursts of a single frequency. As more and more of these types of warnings are added to aircraft the discrimination between the sounds will become more difficult and the chances of confusion between them will increase.

In both the rotary and fixed wing aircraft studied the audio volume of the majority of warnings is not pilot selectable ie. they are presented to the aircrews ears at a fixed level. The level chosen appears to reflect a "better safe than sorry" approach, ie. the sounds are presented at maximum volume and are continuous to guarantee detection. Unfortunately this approach not

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

only contributes to hearing damage risk but causes a startle effect and interferes with communications, usually prompting aircrew to initially seek the audio mute facility rather than dealing with the problem at hand.

Another area for concern is the allocation of sounds to specific warnings. Currently there is no standardisation across aircraft. For example, seven helicopters present a RADALT warning when the limit height on the radio altimeter is transgressed, and, although they are alerting to the same problem, three different attensons are currently being employed. Aircrew who fly regularly in different helicopters or who are converting from one aircraft type to another may find the practice of using different attensons to alert to the same warning confusing. Similarly, the survey showed that the same attenson is being used to alert to different warnings in different aircraft, a practice that could be equally confusing.

In summary, the sounds that have been introduce to aid aircrew are generally ill-considered and have not been designed as an integrated set. Consequently the audio warnings currently in use may prove counter-productive and have flight safety implications which, if allowed to continue, may eventually prove catastrophic.

For the last ten years the DRA (formerly the Royal Aircraft Establishment, RAE) has conducted a programme of research into audio warning design and presentation. The work has shown that existing problems are avoidable. Advances in computer technology now make it possible to produce more complex artificial sounds that can be tailored in terms of frequency content for maximum effect in a given noise environment and software has been developed to calculate the predicted auditory masked threshold for a given noise field which allows precise definition of the levels at which audio warnings should be presented for reliable detection. This paper details the research performed to date and outlines the research that will be addressed to provide a fully integrated audio warning suite for standardised use across the UK military fleet.

2. RESEARCH PERFORMED TO DATE

2.1 Background

Research into the use of audio warning signals in military aircraft began in the early 1980s supported by the Ministry of Defence (MOD) and what was previously the Royal Aircraft Establishment (RAE). The work was performed in conjunction with the Institute of Sound and Vibration Research (ISVR) at Southampton University, the Medical Research Council's (MRC) Applied Psychology unit (APU) at Cambridge and the Department of Psychology at the University of Plymouth. Throughout the research close liaison was maintained with the test pilots at both RAE Farnborough and A&AEE Boscombe Down. The research has culminated in a set of design guidelines based on psychoacoustic and acoustic research and, an auditory warnings implementation strategy.[2,3,4]

The initial approach adopted in setting the guidelines was to minimise the chances of aircrew missing the audio presentation under high workload conditions. Hence, it was decided that audio warnings should consist of a sequence of repeats of a carefully designed attention getting sound (attenson) followed by a voice message. The attenson was to be a unique sound in the cockpit environment, designed to cut through all other cockpit noise to alert the pilot that a problem had arisen. A digitised female voice would then pin-point the exact problem area. It was considered that a voice message alone could easily go undetected either amongst all the other radio communications that exists, or at moments of emergency when audio speech messages may remain uninterpreted.

2.2 Audio Warning Priority Structure

The overall warning philosophy was designed to be as simple as possible and kept within the existing principles of the visual warning system guidelines laid down in Defence Standard 00-970. It was considered that, over and above the type or position of the problem, the relative urgency of the warning was the critical parameter to be conveyed. Hence, a four tier category of warnings was developed:

Priority 1 - Immediate Action

The highest urgency category where immediate action is required to save the aircraft. The response time is considered to be in the order of 2 seconds.

Priority 2 - Immediate Awareness

Aircrew should be made immediately aware of the problem, but immediate action is not generally necessary eg. Engine fire warnings are rarely acted upon immediately. Aircrew usually check by other means (smoke, engine instruments etc.) before initiating the fire suppression procedures. This category covers the red warnings on the central warning panel.

Priority 3 - Awareness

Aircrew should be made aware of the particular problem but may react in their own time. eg. An anti-ice system failure warning, where the pilot should be aware in order to take some action (which might include a modification of the flight parameters) in the longer term. This category generally covers the amber, or yellow, warnings.

Priority 4 - Information/Status

This is the lowest urgency category and may be used where a change in status of the aircraft or an aircraft system needs to be communicated to the aircrew. Due to the minimal urgency it may be that such audio warnings are not always necessary and that visual warnings, in some cases may suffice.

2.3 Attenson Design

Having set the warning priority structure, attensons had to be designed to associate with each category and to meet the following criteria:

1) To be unique sounds in the cockpit noise environment.

2) To be fully discriminable from all other attensons in the set.

3) To convey the correct relative urgency for the associated priority level

4) To be presented at the correct audio level for reliable detection.

Previous research by Patterson[2] from the APU Cambridge recommended that audio warning attensons should be built from pulses of sounds grouped into bursts (figure1), in effect short melodies or tunes, allowing each attenson to be highly distinctive and memorable. By varying the pitch, tempo and rhythm of the pulses, the urgency of the attenson could be matched to the relative priority levels of the warning structure. Taking into account the spectral content and sound levels of the cockpit noise the spectrum and audio level of the attensons could be designed to produce minimal interference with communications and provide sufficient clarity that they would be heard reliably without startle or distraction.

Using Pattersons auditory masking model[5] the masked thresholds were predicted for a host of helicopter spectra (figure 2) measured at the ear in a number of different helicopter types, flying at various speeds and altitudes. By superimposing all the masked threshold spectra a clear indication was obtained of the frequencies where most of the cockpit sound energy was concentrated. The spectral content of the attensons could therefore be chosen to avoid masking by these dominant cockpit frequencies and a single set of attensons could be designed to be acoustically correct for a wide range of helicopter types.

2.4 Audio Warning Presentation Level

The setting of the audio presentation levels was based upon previous research at MRC/APU using Patterson's auditory masking model. The work is described in detail in reference 5, but the essence of the research is shown in figures 3 and 4. Figure 3 shows a typical noise spectrum measured at the ear of aircrew and the associated auditory masked threshold, ie. the level at which a signal must be presented for a 75% chance of detection in that given noise field. Previous research on psychometric models has shown that presenting a signal 15dB above the masked threshold essentially provides a 100% probability of detection. Due to the temporal and spectral variability of noise spectra within and between helicopter types and with helmet fitting, it was recommended that a 10dB band above the 100% detection level should be provided (figure 4), within which detection would be reliable.

2.5 Development of DRAs attenson suite.

Initially ten warning sounds were produced by APU according to the guidelines previously detailed and, as the ability of pilots to memorise and distinguish audio warnings is crucial, experiments were conducted to measure the performance of aircrew in this respect. Whilst the experiments are discussed in more detail in references 6 and 7, a brief summary of the work is provided for discussion.

2.5.1 Confusion experiments

Experiment 1:

The experiments designed to assess the discriminiability of the 10 attensons used a computer controlled self-paced cumulative learning programme. Ten aircrew were presented with the attensons through the telephones of a flight helmet, worn whilst seated in the DRAs helicopter noise simulator set to generate Sea King cabin noise. An attached microprocessor monitored a panel of buttons with which the subjects identified signals and initiated further presentations.

During the first phase of the experiment subjects underwent a learning exercise. Initially, one warning signal was played and identified to the subject. This identification procedure was followed by a test where the signal was then replayed to the subject who identified it by pressing the appropriate button on the keypad. A second warning signal was then introduced and identified, followed by a test where both warning signals were replayed and identified by the subject. Further warning signals were individually introduced and the test repeated until all ten signals had been introduced and correctly identified. A subject was not allowed to proceed to a further test sequence until all signals presented so far had been correctly identified during the same test mode.

Following the learning exercise subjects returned a week later to conduct phase 2, a revision test. Here subjects had one presentation of each signal and their response (correct or incorrect), was recorded. Phases 3 and 4 (repeats of phases 1 and 2) were then conducted.

In order to determine how easy or difficult it was for subjects to learn this particular combination of signals the total errors monitored during the test phases 1 and 3 were calculated across all subjects for each stage. Figure 5 shows the mean errors and the associated standard deviations for each stage. A change in gradient of the curves can be perceived around the 6th and 7th stage (ie. after the sixth or seventh signal had been introduced into the learning sequence) suggesting that although it is possible to learn and retain a set of signals, it may be relatively more easy to acquire the first seven but greater numbers are hindered by a steeper learning curve.

To identify confusions between the ten signals statistical analysis was performed on the data collected during the test phases to determine which signals elicited an incorrect response more often than indicated by chance. Figure 6 shows the confusion matrices for both phases, with the cells where the values have reached significance being underlined. The data indicated that there was some confusion between signals 3&4 and 7&8.

Experiment 2:

Following the development of a set of "higher urgency" attensons (see 2.5.2) a similar confusion experiment was conducted incorporating four of the signals tested previously, with the six most urgent attensons from the new set. The data collected was analysed in the same fashion and figure 7 shows the confusion matrices for phases 1 and 3 of the experiment. There was little evidence of confusion amongst these ten attensons, with just one confusion being significant. Although few errors were exhibited (22 errors in a total of 641 presentations) figure 8 shows that as seen previously there is a marked change in the error rate at the sixth and seventh stages. However, unlike the first set of results the error values then drop back to a continuation of the previous trends for the succeeding stages, possibly indicating that this set of attensons presented fewer problems during the learning process.

2.5.2 Urgency experiments

Following the confusion experiments on the initial set of ten attensons produced by APU it was recognised that more urgent sounding attensons would be required for the Immediate Action category of warnings. DRA supplied the APU with two example attensons from existing experimental aircraft fits that were known to convey extreme urgency. From these signals APU were able to construct a further set of 20 signals which were each provided in two formats. One format conveying a higher urgency than the other, but both having essentially the same temporal and spectral construction. The signals designed to have the lower urgency were designated as the "Initial" signals, and those intended to be more urgent as "Urgent" signals.

The perceived urgency experiment was divided into four trials. The first three were designed to reduce the new attensons to a set of the six most urgent sounding and, although the main interest was in the most urgent sounding signals, all 40 sounds (20 pairs of Initial and Urgent signals) were assessed during the first two trials. The most urgent ten from each of these experiments were then grouped together for a third trial where the most urgent six were identified. These were then joined with two signals previously assessed in the confusion experiments and intended for use against the Immediate Awareness and Awareness categories of warnings and the two examples of high urgency warnings initially provided to APU, for a fourth trial.

The experiments were paired comparison rank-ordering tasks performed by experienced subjects. The subjects were asked to choose which of two, consecutively presented warning signals sounded more urgent. The signals were presented through earphones and identified on a VDU screen as A or B. A keypad allowed the subjects to record the choice made. All activities were computer controlled and the experiments were self-paced in that the system waited for a response after each pair before proceeding to the next. Rankings for each set of signals were produced for individual subjects and across all subjects.

The rankings produced by the four experiments are shown in Tables 1 to 4. The six signals that proved most urgent from trial three showed that the new signals designed by APU were not only internally consistent, ie. all Urgent versions of a particular signal were always ranked higher than the corresponding Initial version but also, with just three exceptions, all Urgent signals were ranked higher than all Initial versions. These points showing support for the design methods.

Trial four showed that the six new sounds specifically designed to have high urgency were ranked higher than the attensons drawn from the original set of ten intended for use against the Immediate Awareness and Awareness categories of warnings. These in turn exhibited the correct levels of relative perceived urgency. The two high urgency example attensons were ranked comparably with the six Immediate Action signals.

2.6 The DRAs attenson suite and presentation philosophy.

As a result of the experiments addressing confusion between attensons and their relative perceived urgencies a set of ten warning attensons have been produced which are considered to be a fully tested baseline set for use across all helicopter types.

The set consists of:

a) Six Priority 1 (Immediate Action) attensons, all exhibiting a high degree of perceived urgency. The philosophy of standardisation requires each Priority 1 warning to have a dedicated attenson which would be specific to that particular problem across all helicopter types. For example, if "Rotor Droop" required immediate action, then the attenson used in a Gazelle would also be used for "Rotor Droop" in Sea King, or any other helicopter. The attenson would then be followed by a voice message detailing the problem. Hence, a pilot could fly any helicopter type and in a high urgency situation still react to the attenson correctly before hearing the voice message.

b) One Priority 2 (Immediate Awareness) warning, to be alerted by just one associated attenson followed by a unique voice message to actually pin-point the individual problem.

c) One Priority 3 (Awareness) warning, covering all cautionary warnings. This is alerted by a single attenson followed by a single voice message of "Caution" or "Master Caution", intended to direct the aircrew to view the Central Warning Panel (CWP) to pin-point the exact problem.

d) One Priority 4 (Information) warning which is the lowest urgency category and intended to convey information or status details. This would use a single attenson, possibly backed up by a voice message depending on the application.

e) A Low Height Warning:- For helicopters the "Low Height" warning is considered a special category warranting a unique dedicated attenson. The height at which this warning triggers is variable and is actually dictated by the sortie profile. Hence the pilot presets a bug height on the radio altimeter which when transgressed will trigger the "Low Height warning". Hence, the final attenson in the set of ten is dedicated to this category.

2.6.1 Warning presentation sequences

Having determined the audio warning philosophy and built a baseline set of attensons, warning sequences were designed (figure 9) paying particular attention to their length and how often they were repeated. The initial, urgent and background bursts were introduced because although it was hoped that aircrew would respond to the first attenson/voice message sequence presented at the initial level, if for some reason he missed the warning or failed to acknowledge it by either rectifying the problem or cancelling the audio, the warning should be presented again in the appropriate time scale and at a level reflecting the urgency of the priority level of the warning. For the priority 1 warnings, as the response time is considered to be in the order of 2 seconds only, a short sequence consisting of two presentations of the Priority 1 attenson and the associated voice message was designed to be presented at the urgent level straight away. The sequences shown were designed for flight assessment in the DRA's Sea King helicopter.

The initial aircrew reaction was generally against the long sequences on the basis that the warning had been acquired by the aircrew early in the sequence initiation and that any further messages were superfluous. Whilst this may be so for low workload situations no research has been carried out under high workload or high stress conditions. Such research is difficult since, even in aircraft simulators, high stress/workload situations are notoriously difficult to reproduce with any fidelity. In the final instance, it is possible to crash a simulator without adverse effect on the aircrew. By definition high workload or high stress in an aircraft during flight trial evaluation means higher risk to aircraft safety and thus experimentation is often not acceptable.

Having considered the Sea King pilot's comments and taking into account the requirements for EH101, it was decided to approach the problem from the other end and provide a set of sequences considered to be a minimum acceptable for UK use. A new set of shortened sequences were designed, adopting just one presentation of the attenson and voice message at all priority levels. However, the "Low Height" warning was maintained as a continuous sequence, where continual reminders until correction were considered vital.

These shortened sequences were installed in the DRA's experimental Lynx aircraft where over a period of time a number of test pilots routinely flew with the warning system. The subjective opinions of the aircrew are presented in detail in reference 8 but generally, the consensus of opinion was that this type of audio warning suite has a place in the next generation of aircraft and, if a standardised set can be agreed upon, it would be beneficial to start retrospective fitting.

2.7 Varying Aircraft Parameters

The survey conducted by DRA of audio warnings currently installed in military aircraft showed that one of the major groups of audio warnings were those related to variable aircraft parameters. These types of warnings are indications of how a flight parameter, such as rotor overspeed, "g", torque, bank, pitch etc. deviate from a normal operating level towards the extremes of the operating envelope. They are in fact trend indicating sounds (trendsons) as opposed to warnings and a rudimentary attempt was made to provide indications for these varying parameters during the flight trials in the DRA's experimental Lynx helicopter[8]. These trials showed that as for the aircraft system warnings, trendsons are a distinct warning set requiring their own presentation strategy. Under an MOD contract, the Psychology Department at the University of Plymouth defined the characteristics a trendson should exhibit and investigated the information they should convey. The research culminated in the development of a protocol for trendson design and is detailed in references 9 and 10. However, the essence of the work is summarised in the next sections.

2.7.1 Trend Indicating Sounds (Trendsons)

Trendsons are intended to provide feedback to the pilot when the particular aircraft parameter being conveyed has begun to exceed normal limits. As normal limits are further exceeded, the characteristics of the trendson should alter in such a way as to convey the direction of the change and the speed with which the parameter is changing. If the critical point is exceeded then additional information should be supplied in the form of an audio warning.

The most effective way of conveying change through sound was shown to be by the use of different levels of very short, discrete units of sound. As the time histories of events that would be conveyed through trendsons are relatively short (the events can range from just above normal limits when the trendson should come on, to critical limits when a warning

should sound, in a matter of seconds) it was recommended that a trendson should consist of five separate levels at most. Each level consists of a unit of sound which would be triggered once a preset value of the parameter being conveyed, eg. rotor overspeed, is exceeded. This level should continue playing until either the speed falls back to within normal limits or increases such that a second preset value is reached, at which point the second level of the trendson would play in the same way. This sequence of events would progress until the 5th level is exceeded, when a warning should be heard. As the parameter returns to within normal limits, the five levels are heard in the opposite direction until no further sounds are heard. An example of how a trendson might function is shown in figure 10.

For the trendsons to be designed in the most psychologically appropriate way, an extensive series of laboratory studies elucidated the main psychological correlates of many of the acoustic parameters available for use in the design of trendsons. The most important meanings of parameters such as pitch, speed, rhythm etc. were isolated and used to produce an initial set of trendsons which convey change through several acoustic parameters.

From the testing of this initial set it became apparent that acoustic changes convey a variety of meanings. Some parameters used in the design of the trendsons conveyed different meanings to the listener and, on occasions, these meanings reinforced one another whilst at others were contradictory. For example, a falling, slowing pitch pattern not only conveyed an object falling and/or slowing down, but also that a situation was becoming less urgent. If such a pattern was used to convey rotor underspeed, clearly, contradictory information would be presented.

Consequently, further research addressed the meanings conveyed by different acoustic parameters with a view to minimising the effects of contradictory information. The meanings associated with the most important acoustic features of the trendsons set were quantified, allowing the relative strengths of each meaning to be assessed. The information allowed the most compelling meaning associated with each trendson to be conveyed whilst minimising the effects of undesirable meanings.

Based on this research a set of five trendsons have been produced, one each for rotor overspeed, rotor underspeed, power, positive "g" and negative "g". All five trendsons consist of five discrete levels which are acoustically related but different. The acoustic and temporal characteristics of each trendson are distinguishable from one another and should be easily learnt. The direction of the trend is conveyed by the acoustic changes at each level and the nature of the particular parameter being conveyed is, to a greater or lesser extent, implicit within the trendson itself, which should reduce learning time.

When implemented the parameter values at which each level of the trendson should trigger must be predetermined. If all five levels of an individual trendson are not required, to preserve the identity of the trendson, consecutive subsets should chosen ie. levels 1,2 & 3, levels 2,3 & 4 or levels 3,4 & 5 and not for example, levels 1,3 & 5.

2.8 Conclusion of the research to date

When DRA embarked on research into use of audio warnings in the military cockpit there were three main areas that needed to be addressed.

Firstly, the numbers of warnings being installed:- The research has indicated that the maximum number of sounds that aircrew can easily learn and retain as having specific meanings is about seven. Consequently the aim should always be to minimise the number of attensons in an aircrafts warning set to less than seven. This may be achieved by adopting the prioritised categories of warnings philosophy where the number of attensons can be limited to between four and ten, depending on the number of Immediate Action Warnings required. In general very few aircraft specify audio requirements for priority one warnings. However, those that do may require only one or two dedicated attensons. Hence, this approach limits the number of sounds in the warning suite but remains flexible enough to allow new warnings to be added at a later date without necessarily increasing the number of attensons in the set.

Secondly, the types of sounds being installed:- The sounds traditionally used as warning signals have been simply constructed in the form of alternating tones, frequency sweeps and repetitive bursts of single frequencies. This type of attenson is easily masked by dominant cockpit frequencies and can be easily confused with warnings with similar frequency content. Similarly, warnings with similar characteristics may also be confused, eg. two frequency sweeps, albeit over different frequency ranges, may under high workload be detected simply as a sweep and the frequency content be indistinguishable. By adopting the more complex warning signals built from pulses of sounds, the attensons can be specifically tailored not only for the noise environment they will be presented in but to be highly discriminable and to have the correct levels of relative perceived urgency.

Thirdly, presentation levels of the sounds installed:- The survey conducted of audio warnings currently used in military cockpits showed that in the majority of cases the warnings were being presented at a fixed volume, directly to the aircrews ears. The warnings are generally presented on a "better safe than sorry" basis ie. too loudly, to guarantee detection. This practice can be counter-productive in that aircrew avoid procedures that knowingly trigger the sounds or are so startled when a sound is presented that the initial reaction is to cancel the audio rather than dealing with

the problem at hand. Such problems are avoidable, a computer model now exists that can accurately predict the level warnings should be presented at in a given noise environment to allow reliable detection and this not only eliminates startle effects but also reduces hearing damage risk.

The survey of audio warnings currently used in military aircraft showed three distinct groups of warnings requiring audio backup, namely Aircraft System Failures, Variable Aircraft Parameters and Aircraft Threats. To date the research has culminated in the production of a set of audio warning design guidelines that have been used to produce a suite of attensons, purpose built for standardised use across the UK military helicopter fleet. These attensons and the implementation strategy have been adopted for presenting aircraft failure and threat warnings in Merlin, the Naval variant of the EH101 helicopter. Design guidelines and an implementation strategy have also been addressed for variable aircraft parameters and a set of five trendsons have been produced, although have yet to be test flown. However, in an effort to enhance the ability of aircrew to manage and process a greater number of sounds DRA have been looking for new techniques for presenting warning sounds. One technique that appears promising monopolises on the human's ability to accurately localise sounds in the environment. It is possible that by presenting dedicated threat related attensons in three dimensional space at the apparent location of a threat, aircrew may be able to respond more quickly and accurately to the warning. As the threat warnings would be spatially separated from the aircraft system warnings there is potential to learn/recognise and react correctly to a greater number of attensons.

The following sections discuss the programme of work required to assess the feasibility of localising aircraft threat warnings in the cockpit and the issues that need to be addressed if all three categories of warnings (Aircraft System Failures, Variable Aircraft Parameters and Aircraft Threats) are to be fully integrated into a complete warning suite.

3. DESIGN CONSIDERATIONS OF AUDIO WARNINGS FOR SPATIAL LOCALISATION.

As discussed in the previous section there is a growing interest in utilising the humans' localisation abilities to detect aircraft threat related warnings presented in 3D auditory displays. The main advantage of adopting this type of presentation philosophy is that it provides a map of auditory space and can immediately alert aircrew to the location of the threat, potentially resulting in both quicker and more accurate reactions, which may prove crucial at times of emergency.

3.1 Synthesis of sounds in 3D space

To accurately simulate 3D sound it is necessary to successfully synthesise the Binaural, Monaural and Positional cues that enable us to locate. If these cues can be encoded in a signal presented to aircrew through the communications telephones, the signal would appear to originate from it's designated location.

There now exist a number of commercially available devices that utilise current knowledge of localisation cues to synthesise sound localisation. To encode the effects of the head, torso and pinna, recordings are made of an individuals Head Related Transfer Function (HRTF). This is achieved by inserting small probe microphones into the ear canal and recording the impulse response of a sound source positioned at a number of locations about the head. The recordings made contain all the modifications made to the signal by the head, torso and pinna for a particular location. The recordings from both ears for each position are analysed and used to create digital filters which produce the same phase and amplitude effects as the head, torso and pinna. By filtering a signal through these filters the sound can be made to appear as though it originated from the initial recorded location ie. the signal is spatially encoded. As it is impossible to record at all positions on a sphere about the head, linear interpolation is used to generate the filters for those locations between the recorded positions.

Some devices have also integrated the output of the system with a head tracker making it possible to synthesise the effects of the relative positions of the source and the observer. That is, if a signal is initially presented behind a listener who then turns his head 90^0 to the right, the tracker will detect the head movement and the output from the synthesiser will alter so the listener will then locate the signal as coming from the right.

Whilst the sound localisation synthesisers can work well for many listeners they are still unrefined and may exhibit front-back reversals, an inability to perceive external sound from the headphones and lack perception of elevation. Previous experience of using these systems has shown that the strongest sense of "acoustic reality" and the greatest accuracy of localisation is achieved when the listener uses his own HRTF. Also, systems that use a greater number of source locations in the recording of the HRTF appear more refined, giving better definition.

Whilst ideally is would be best for all listeners to have their own HRTF encoded in the synthesiser, it takes some two hours to record and is therefore a costly and time consuming exercise. Hence, future research will

need to address the HRTF issue, looking at the feasibility of adopting a "sensible average" and whether it would provide enough sensory cues for the majority of listeners to pinpoint sounds with a reasonable degree of accuracy. The possibility of using of a dummy head fitted with an average pinna for recording the HRTF should also be investigated, as should the possibility of exaggerating the location cues. It may be that by magnifying the time and intensity differences, as if the ears had been moved further apart, better location resolution may be achieved.

3.2 Sound parameters required for accurate localisation.

Signals with different frequency content provide different types of localisation cues. Research shows that the monaural cues required for the discrimination of front/back and elevation are derived from modifications to the high frequency end of the signal spectrum (>4kHz). This is because the wavelength of the sound has to be short to interact with the pinna and suggests that the high frequency content of a signal is important if it is to be localised accurately. Furthermore, it has been suggested that the cues for different locations are frequency specific. Consequently, if the frequency cues for a particular location are absent from a signal the listener will perceive the location of the sound as corresponding to the frequencies that are present, irrespective of the actual location of the sound source. This is supported by experimental findings that show broadband, white noise is located better than any other sounds, implying that the wide range of frequencies the human auditory system uses for localisation is adequately represented in broadband, white noise.

DRA's previous research on audio warning design has concentrated on providing sounds that can be heard reliably over the flight helmet telephones without being masked by the high noise levels present at the aircrews' ears. As detailed in section 2.3, the spectral content of the pulses of sounds used to build the audio warnings, was specifically chosen to avoid dominant cockpit frequencies and provided enough spectral redundancy such that the sound would be interpreted as the same sound in different helicopter noise environments. The frequency bandwidth was limited to between 200Hz and 4kHz, the bandwidth of current aircraft communication systems.

Unfortunately, at the time the design guidelines were being set for these types of sounds the characteristics for optimising their localisation cues was not a consideration. Hence, initially the suitability of the current style of audio warnings for use in 3D auditory displays needs to be addressed. Aspects such as frequency bandwidth, spectral content and spectral density have to be investigated and should hopefully culminate in a set of design guidelines for optimising the parameters of a sound such that it can be easily localised in 3D space.

3.3 Presentation of 3D sounds in the cockpit environment.

3.3.1 Cockpit application issues

Although the laboratory may provide a satisfactory environment for assessing the sound characteristics required for optimal localisation cues, when an optimised sound is actually played in the aircraft cockpit a number of environmental conditions may effect their ability to be localised.

Despite an ongoing programme to reduce noise levels at aircrews' ears, high levels of cockpit noise are still transmitted through the flight helmet earshells. Whilst the helmet provides good high frequency attenuation there is relatively little protection at low frequencies[11] and consequently, high noise levels existing at the ear may mask portions of audio warnings presented over the communications telephones. DRA's future work programmes looking at the design of aircraft threat warnings for presentation in 3D auditory displays will address the masking effects of cockpit noise and will investigate the associated effects of introducing Active Noise Reduction (ANR) systems to improve low frequency helmet attenuation.

Another limitation to presenting 3D auditory displays in current aircraft is the bandwidth of the communications system. As previously discussed, different localisation cues are frequency specific. Future research will establish which localisation cues are most dominant and will therefore reveal the frequency bandwidth required for accurate localisation. Undoubtedly, future specifications for aircraft communications systems will require an extended bandwidth from the existing 200Hz to 4kHz to possibly, 100Hz to 8kHz. Fortunately other aspects of audio communications such as Active Noise Reduction systems require higher quality and wider band transducers and thus support the requirement of the wider band communication systems necessary for spatial localisation. Also, if lightweight flight helmets incorporating ear insert style communication devices are specified, the effects on localisation of presenting sounds nearer to the eardrum will have to be taken into consideration and consequently future research will investigate this aspect.

3.3.2 Presentation philosophy.

For the DRA's existing audio warning suite the presentation philosophy requires an attention getting sound to alert the aircrew to the existence and priority of a problem that has arisen and then a follow up voice message to pinpoint the exact details of the problem. For aircraft threat warnings, however, three different pieces of information need to be conveyed to the pilot.

i) threat status

ii) threat location

iii) threat type

Through collaboration with the US Army under the auspices of TTCP-HTP6, DRA have categorised the threat types into two groups, Radar (missiles, guns and unknowns) and Laser, and the threat status levels into:

i) Search

ii) Acquisition

iii) Track

iv) Launch

However, further discussions with aircrew will be required to confirm all threat types and status levels that need to be considered.

Whilst presenting a warning sound in 3D space provides positional information, future research will address how the threat status and type should be conveyed. It could be that if a radar is just searching no audio would be required and the visual displays would be sufficient, or possibly, a sound with low perceived urgency could be presented at a low audio level localised in the radars direction. As the status of the threat changes to a higher level the attenson may get louder or more urgent sounding. Previous work by DRA has shown that the parameters of a warning attenson can be varied such that the perceived urgency of the sound will vary but the essence of the sound remains the same, ie. the attenson is still recognised as being the same sound. This would enable different threats and their status levels to be depicted by an individual attenson and their location mapped in 3D space. However, the overriding philosophy should always aim to reduce the number of attensons in the set to a minimum in order to maintain the aircrews' ability to learn and discriminate.

Simulation will show whether adequate information can be provided via spatially located attensons. It may be that backup to the audio may be required in the form of a voice message and hence, voice message construction will be addressed in future research. Simulation will also show whether refined localisation of warnings in 3D space is actually necessary. Whilst humans can localise to small degrees of accuracy and technology may be capable of presenting sounds at fine positional resolution, it is possible that with well designed EW visual displays sufficient information can be provided to the aircrew by simply presenting the sound in the relevant quadrant. For the added expense of more sophisticated technology the extra resolution may provide no added advantage in terms of aircrew reaction time. Hence, the development of the threat related audio displays will be closely matched to the evolving designs of the visual displays.

4. CONCLUDING DISCUSSION

Whilst every pilot has his own opinion of what he considers a good attenson and what warnings he feels should be covered by the audio system it is not possible to cover every combination. As the number of audio warnings increase the potential for mis-interpretation and confusion is increased. The practice of individually designing warnings in isolation from others within and between aircraft types will inevitably have flight safety implications with possible catastrophic consequences. Research to date has provided a structure in which the two major categories of audio warnings (Aircraft System Failures and Aircraft Threats) can be presented. It employs a minimal number of attensons (all easily recognisable and discriminable) which can be presented at levels that will allow reliable detection without being intrusive and, although initially tailored for helicopters would, with slight modifications, be suitable for fixed wing aircraft. A set of design guidelines have also been produced for the third major set of alerts, Variable Aircraft Parameters. This research, integrated with the work to be conducted in the near future on spatially located threat warnings will mean that by taking the audio warning requirements for an aircraft as a whole it should be possible to provide a well balanced, fully integrated warning system within a framework for standardisation across the UK military aircraft fleet.

5. REFERENCES

1 Susan H Mungur, "A survey of audio warnings currently used in military aircraft and an assessment of their suitability for operational use." Unpublished DRA Tech. Report, 1993.

2 R. Patterson "Guidelines for auditory warning systems in civil aircraft." CAA Paper 82017, 1982.

3 M. Lower, P. Wheeler, R. Patterson, Judy Edworthy, G. Rood and J. Chillery, "The design and production of audio warnings for helicopters. Vol.1, SeaKing." Unpublished contract report by ISVR, 1986.

4 Judy Edworthy, Sarah Loxley, E. Geelhoed and I. Dennis, "An experimental investigation into the effects of spectral, temporal and musical parameters on the perceived urgency of audio warnings." Unpublished contract report by Plymouth Polytechnic, 1988.

5 R. Patterson, "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram and speech threshold." JASA, 72, 1788-1803 (1982)

6 J. Chillery and J. Collister, "Possible confusions amongst a revised set of auditory warning signals developed for helicopters." Unpublished RAE Tech. Memo. 1986.

7 J. Chillery and J. Collister, "Perceived urgency of warning signals determined using a forced choice pair comparison technique." Unpublished RAE Tech. Memo. 1988.

8 Susan H James, "The development of an integrated warning set for military helicopters and the initial flight trials in Lynx ZD285." Unpublished RAE Tech. Memo.1990.

9 Judy Edworthy, Sarah Loxley and Elizabeth Hellier, "A preliminary investigation into the use of sound parameters in the portrayal of helicopter trend information." Unpublished contract report by Plymouth Polytechnic, 1989.

10 Judy Edworthy, Sarah Loxley and Elizabeth Hellier, "Trend Audio Information Signals" Unpublished contract report by Plymouth Polytechnic, 1992.

11 M. Hancock and Lysanne Rood, "A comparative assessment of the acoustic noise attenuation provided by the Mk4A, Mk4B and Mk10B aircrew helmets." Unpublished DRA Working Paper, 1994.

TABLE 1

SIGNALS IN TRIAL A RANKED IN ORDER OF URGENCY

Ranked signals in first set of 20 (over 6 SS)

Score CL	Signal	Score GR	Signal	Score RW	Signal	Score DH	Signal	Score SL	Signal	Score JB	Signal	Score All	Signal
94.7	20	97.3	10	81.5	20	100.0	12	89.4	12	97.3	12	91.2	12
86.8	12	92.1	12	81.5	12	92.1	5	84.2	10	86.8	14	81.5	10
86.8	10	89.4	20	73.6	14	76.3	13	78.9	9	86.8	5	72.8	20
73.6	18	76.3	13	73.6	10	76.3	4	71.0	7	73.6	10	67.9	4
71.0	19	71.0	4	71.0	4	73.6	10	71.0	6	68.4	20	67.1	5
68.4	8	68.4	17	65.7	19	68.4	20	68.4	14	65.7	13	62.2	14
65.7	4	65.7	5	63.1	8	68.4	17	60.5	4	63.1	4	60.5	19
60.5	5	63.1	8	63.1	5	63.1	19	57.8	11	57.8	19	59.2	17
57.8	17	57.8	19	60.5	18	57.8	14	52.6	17	52.6	7	57.4	13
52.6	7	55.2	11	57.8	17	55.2	11	52.6	15	50.0	18	57.0	7
50.0	14	55.2	7	57.8	7	52.6	7	52.6	2	50.0	17	50.4	8
39.4	15	47.3	18	50.0	13	50.0	2	47.3	19	47.3	11	50.0	18
39.4	13	42.1	9	44.7	9	36.8	18	42.1	8	44.7	9	47.8	11
39.4	11	36.8	14	39.4	15	34.2	9	36.8	13	39.4	8	47.3	9
39.4	9	26.3	15	36.8	16	28.9	6	34.2	20	34.2	15	35.9	15
28.9	16	23.6	6	31.5	11	26.3	8	34.2	5	31.5	2	31.5	2
23.6	2	15.7	2	18.4	6	23.6	15	31.5	18	21.0	6	29.8	6
15.7	6	10.5	1	15.7	2	7.8	16	18.4	1	13.1	16	16.6	16
5.2	3	2.6	16	10.5	3	7.8	1	10.5	16	10.5	3	7.4	1
0.0	1	2.6	3	2.6	1	0.0	3	5.2	3	5.2	1	5.7	3

TABLE 2

SIGNALS IN TRIAL B RANKED IN ORDER OF URGENCY

Ranked signals in SECOND set of 20 (over 6 SS) (21-40)

Score SL	Signal	Score CL	Signal	Score RW	Signal	Score DH	Signal	Score GR	Signal	Score JB	Signal	Score All	Signal
92.1	33	89.4	33	86.8	30	100.0	25	100.0	30	86.8	34	85.9	30
92.1	30	86.8	40	78.9	40	86.8	33	92.1	32	84.2	25	78.5	33
78.9	27	86.8	39	76.3	39	84.2	32	81.5	39	81.5	30	76.7	39
71.0	29	81.5	35	76.3	34	78.9	39	81.5	37	78.9	32	75.8	32
65.7	31	78.9	30	73.6	32	78.9	35	76.3	40	73.6	39	72.3	25
63.1	39	76.3	37	71.0	33	76.3	30	76.3	24	71.0	40	71.9	40
63.1	37	63.1	32	71.0	25	71.0	40	68.4	33	63.1	33	64.9	37
63.1	32	63.1	25	68.4	35	65.7	24	68.4	25	63.1	31	59.6	35
57.8	35	57.8	34	63.1	24	60.5	37	57.8	27	57.8	24	59.6	24
50.0	26	55.2	38	57.8	37	52.6	34	52.6	28	50.0	37	56.1	34
47.3	40	47.3	27	55.2	38	47.3	22	50.0	31	47.3	28	50.8	31
47.3	25	47.3	24	47.3	28	44.7	31	39.4	29	47.3	27	50.8	17
47.3	24	44.7	31	36.8	31	36.8	27	36.8	34	39.4	26	39.4	28
44.7	22	28.9	28	36.8	27	31.5	26	34.2	35	36.8	35	37.2	38
34.2	28	26.3	29	28.9	29	26.3	38	31.5	38	34.2	22	33.7	29
26.3	38	26.3	26	18.4	26	26.3	28	18.4	26	28.9	38	30.7	26
26.3	34	23.6	22	15.7	36	15.7	29	18.4	22	21.0	29	30.2	22
13.1	36	7.8	21	15.7	21	10.5	21	10.5	21	21.0	21	12.7	21
10.5	21	5.2	36	13.1	22	2.6	36	5.2	23	10.5	23	6.5	36
5.2	23	2.6	23	7.8	23	2.6	23	0.0	36	2.6	36	5.7	23

TABLE 3

SIGNALS IN TRIAL C RANKED IN ORDER OF URGENCY

Ranked signals in third set of 20 (over 2 SS)

Score RW	Signal	Score JB		Score All	Signal
92.1	40	94.7	30	82.8	30
76.3	33	89.4	35	80.2	40
73.6	20	76.3	12	76.3	25
73.6	10	73.6	32	67.1	33
71.0	30	68.4	40	63.1	34
68.4	34	68.4	5	59.2	32
63.1	25	57.8	39	59.2	10
57.8	39	57.8	34	57.8	39
55.2	35	57.8	33	57.8	12
52.6	37	44.7	37	52.6	20
44.7	32	44.7	10	48.6	37
42.1	19	42.1	19	48.6	5
39.4	24	42.1	13	42.1	19
39.4	12	34.2	14	38.1	35
36.8	14	31.5	20	35.5	14
31.5	7	28.9	24	34.2	24
28.9	5	23.6	7	27.6	7
23.6	17	23.6	4	25.0	13
21.0	4	21.0	35	22.3	4
7.8	13	18.4	17	21.0	17

TABLE 4

SIGNALS IN TRIAL D RANKED IN ORDER OF URGENCY

Ranked signals in FOURTH set (of 10 over 5 SS)

Score RW	Signal	Score DH	Signal	Score TW	Signal	Score JB	Signal	Score KH	Signal	Score All	Signal
100.0	MLB	94.4	33	88.8	MLB	94.4	30	94.4	LB	81.1	MLB
88.8	40	77.7	MLB	88.8	33	88.8	A	88.8	MLB	65.5	30
77.7	LB	77.7	25	77.7	LB	66.6	32	77.7	30	63.3	33
61.1	30	61.1	32	66.6	40	50.0	MLB	61.1	33	60.0	LB
44.4	33	61.1	40	55.5	34	50.0	34	55.5	25	57.7	40
44.4	32	50.0	34	55.5	30	38.8	40	50.0	32	48.8	32
33.3	34	38.8	30	27.7	25	33.3	25	33.3	40	45.5	25
33.3	25	27.7	LB	22.2	32	27.7	IA	22.2	IA	41.1	34
16.6	IA	11.1	IA	16.6	IA	27.7	33	16.6	34	18.8	IA
0.0	A	0.0	A	0.0	A	22.2	LB	0.0	A	17.7	A

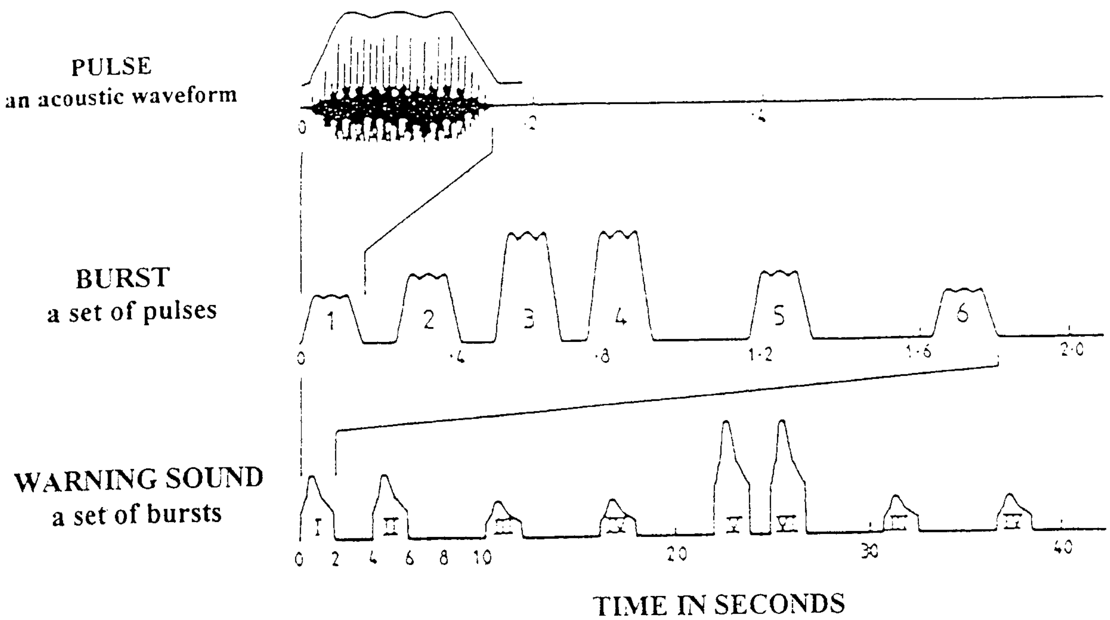


Figure 1 *The building blocks for an audio warning.*

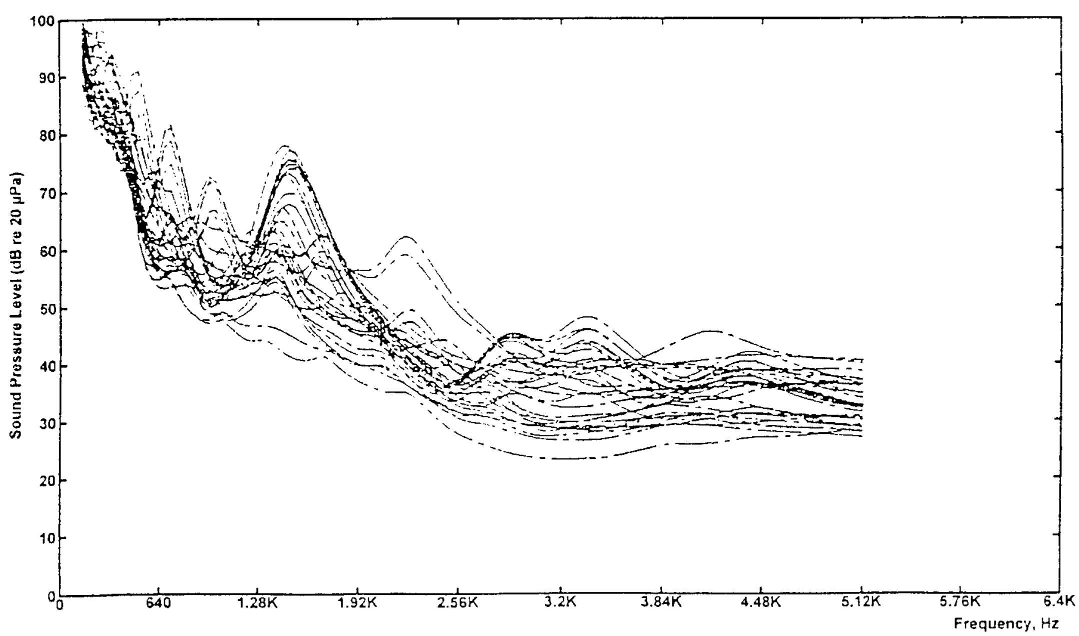


Figure 2 *Predicted auditory masked thresholds for helicopters at different speeds and heights.*

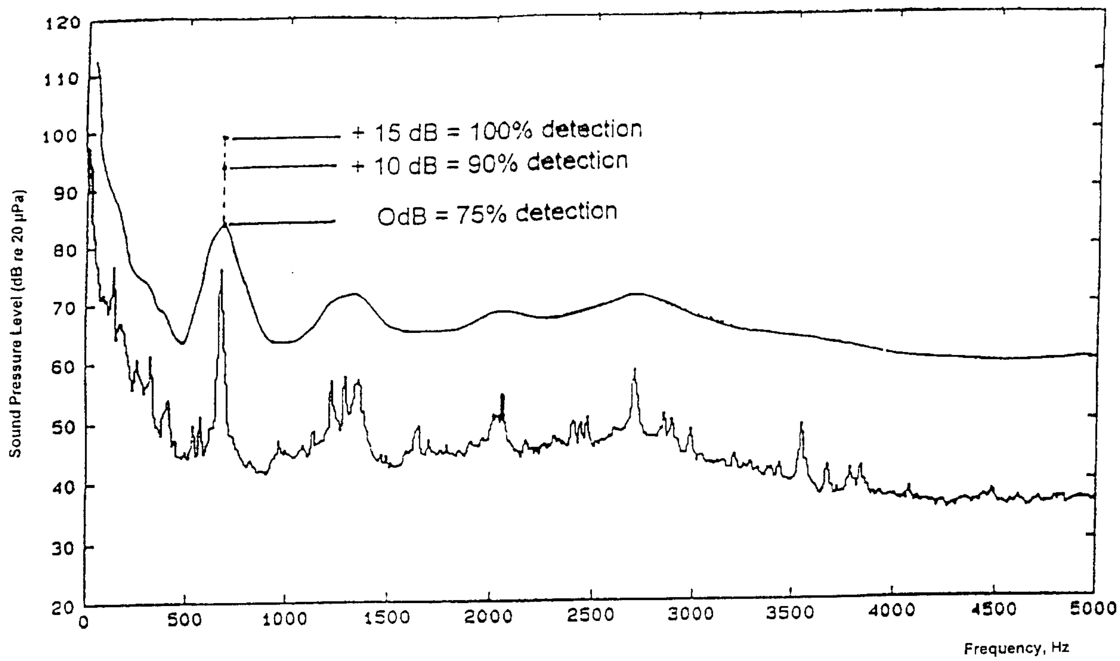


Figure 3 *Noise levels at the ear and the predicted auditory masked threshold.*

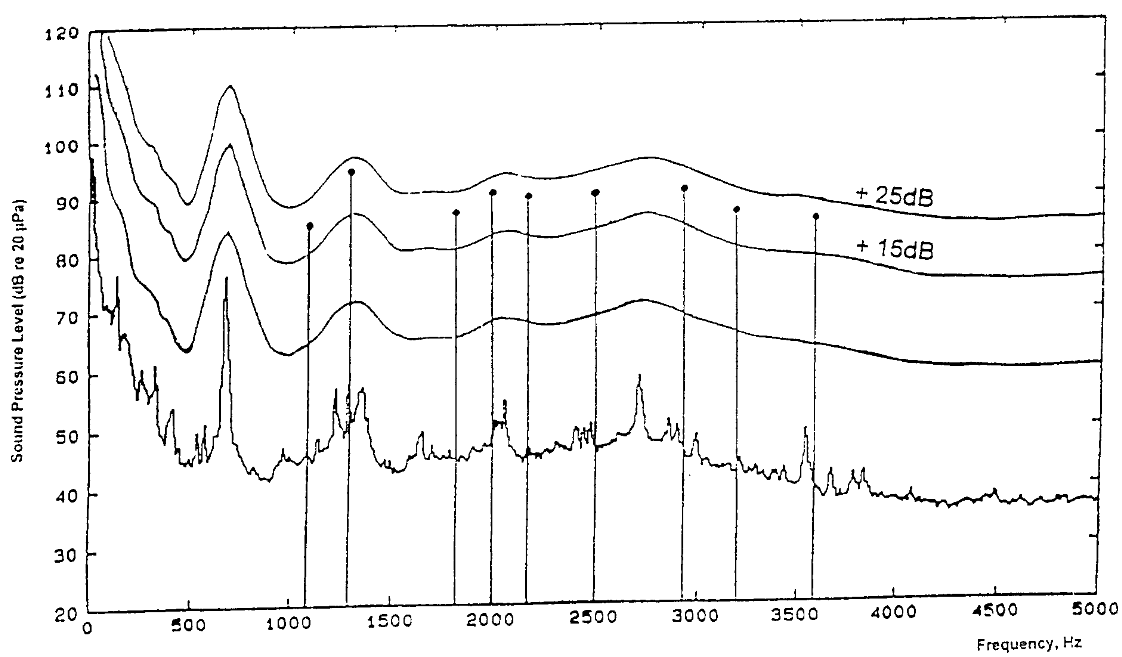


Figure 4 *The 10dB band above masked threshold for reliable detection.*

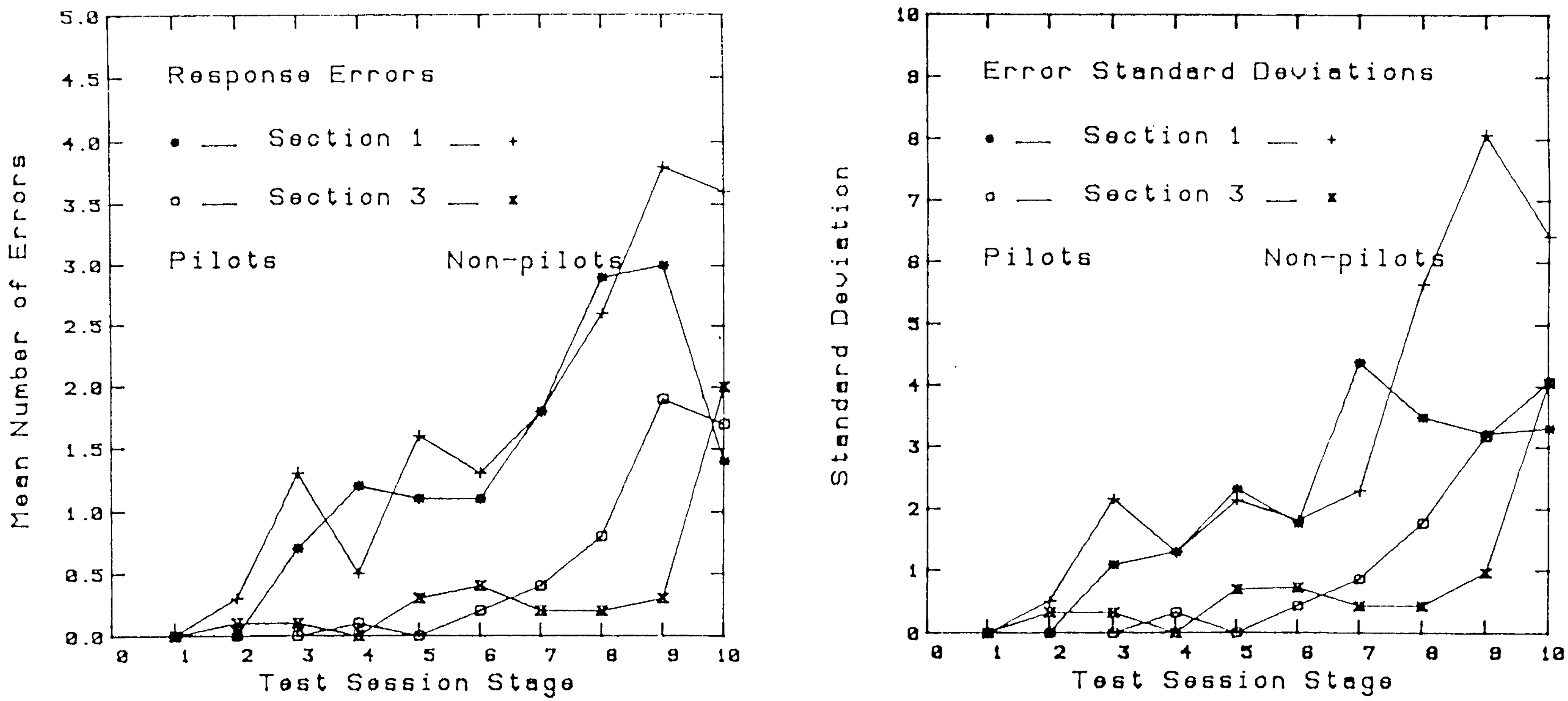


Figure 5 *Mean number of errors during the test stages of confusion experiment 1.*

Confusion Table

Section 1 - Pilots

Signal Presentation	Responses 1	2	3	4	5	6	7	8	9	10	TOTAL
1 FIRE	88	1	5	3	0	0	1	1	0	0	99
2 SERVO	0	110	2	3	0	4	0	2	0	5	126
3 TRANSMISSION	3	4	89	11	0	0	2	0	1	0	110
4 GEARBOX	3	2	7	90	0	1	0	5	0	0	108
5 THREAT	0	2	2	0	106	0	2	4	0	0	116
6 ELECTRICS	1	5	2	2	0	105	4	5	3	2	129
7 MASTER CAUTION	0	0	1	1	2	4	97	7	0	0	112
8 PRIORITY 2	1	0	1	0	0	2	6	89	2	0	101
9 PRIORITY 3	0	0	0	0	0	0	1	1	113	0	115
10 LOW HEIGHT	0	2	0	0	0	1	0	0	0	113	116
	96	126	109	110	108	117	113	114	119	120	

Confusion Table

Section 3 - Pilots

Signal Presentation	Responses 1	2	3	4	5	6	7	8	9	10	TOTAL
1 FIRE	69	2	1	1	0	1	0	1	0	0	75
2 SERVO	0	81	0	0	0	1	0	0	0	0	82
3 TRANSMISSION	0	0	73	4	0	0	4	1	1	0	83
4 GEARBOX	1	0	5	72	0	0	4	1	0	0	83
5 THREAT	0	0	0	0	70	0	1	1	0	0	72
6 ELECTRICS	0	1	0	0	0	81	0	0	0	1	83
7 MASTER CAUTION	1	0	2	1	0	2	69	2	0	0	77
8 PRIORITY 2	0	0	0	0	0	2	4	74	3	0	83
9 PRIORITY 3	0	0	0	0	0	0	0	1	76	0	77
10 LOW HEIGHT	1	0	0	0	0	0	0	0	0	78	79
	72	84	81	78	70	87	82	81	80	79	

Figure 6 *Confusion matrices for the ten warnings tested during confusion experiment 1.*

Responses for Section 1

SIGNAL PRESENTATION	1	2	3	4	5	6	7	8	9	10	TOTAL
1 FIRE	90	0	2	1	1	0	1	0	0	0	95
2 ELECTRICS	4	89	0	1	0	0	4	0	0	0	98
3 INFORMATION	2	1	94	3	0	0	0	0	0	0	100
4 LOW HEIGHT	0	1	0	91	1	0	3	1	0	0	97
5 THREAT	0	1	0	0	87	2	1	2	3	2	98
6 UNDER CARRIAGE	0	0	0	0	0	97	1	1	0	3	102
7 FUEL	2	2	1	0	4	2	78	1	7	5	102
8 SERVO	2	2	0	1	4	1	3	78	3	2	96
9 ROTOR	0	0	0	1	1	4	4	0	85	7	102
10 GEARBOX	2	1	1	0	5	3	1	2	5	88	108
	102	97	98	98	103	109	96	85	103	107	

Responses for Section 3

SIGNAL PRESENTATION	1	2	3	4	5	6	7	8	9	10	TOTAL
1 FIRE	60	1	0	0	0	0	1	0	1	0	63
2 ELECTRICS	1	66	0	0	0	0	0	0	0	0	67
3 INFORMATION	0	2	56	0	0	0	0	0	0	0	58
4 LOW HEIGHT	0	0	0	66	0	0	0	0	0	0	66
5 THREAT	0	0	0	0	58	0	0	1	0	0	59
6 UNDER CARRIAGE	0	0	0	0	0	67	0	0	1	0	68
7 FUEL	0	0	0	0	0	0	64	0	2	2	68
8 SERVO	0	0	0	0	0	0	0	62	0	0	62
9 ROTOR	0	0	0	0	0	1	1	0	57	2	61
10 GEARBOX	0	0	0	0	1	0	4	0	1	63	69
	61	69	56	66	59	68	70	63	62	67	

Figure 7 *Confusion matrices for the warnings tested during confusion experiment 2.*

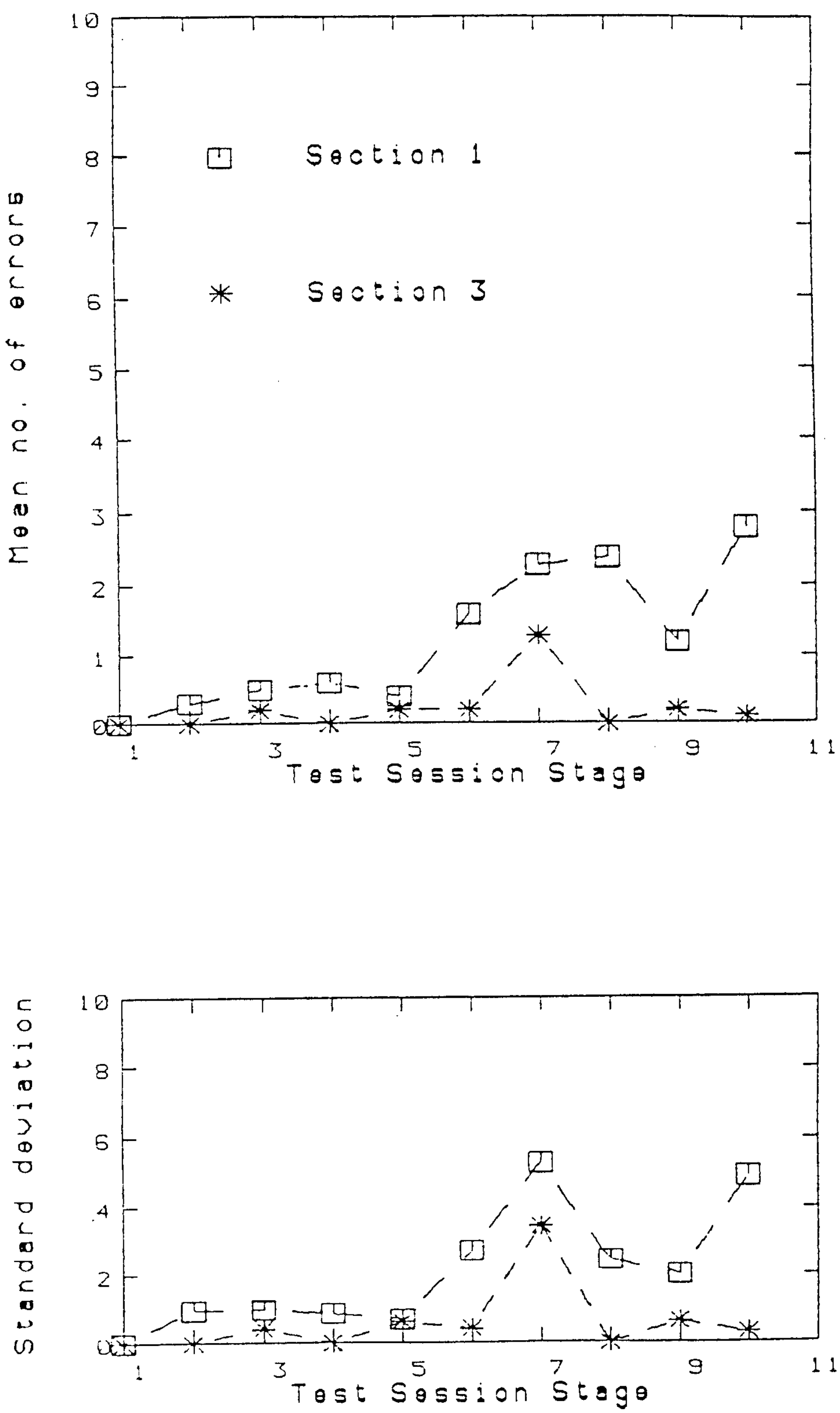


Figure 8 *Mean number of errors during the test stages of confusion experiment 2.*

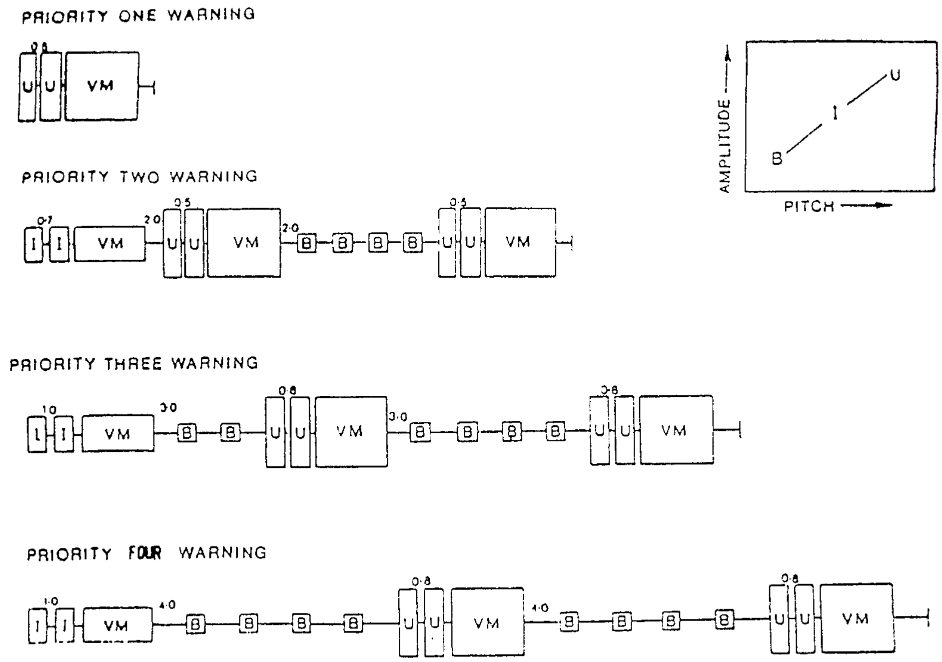


Figure 9 *The original long warning sequences designed for Sea King.*

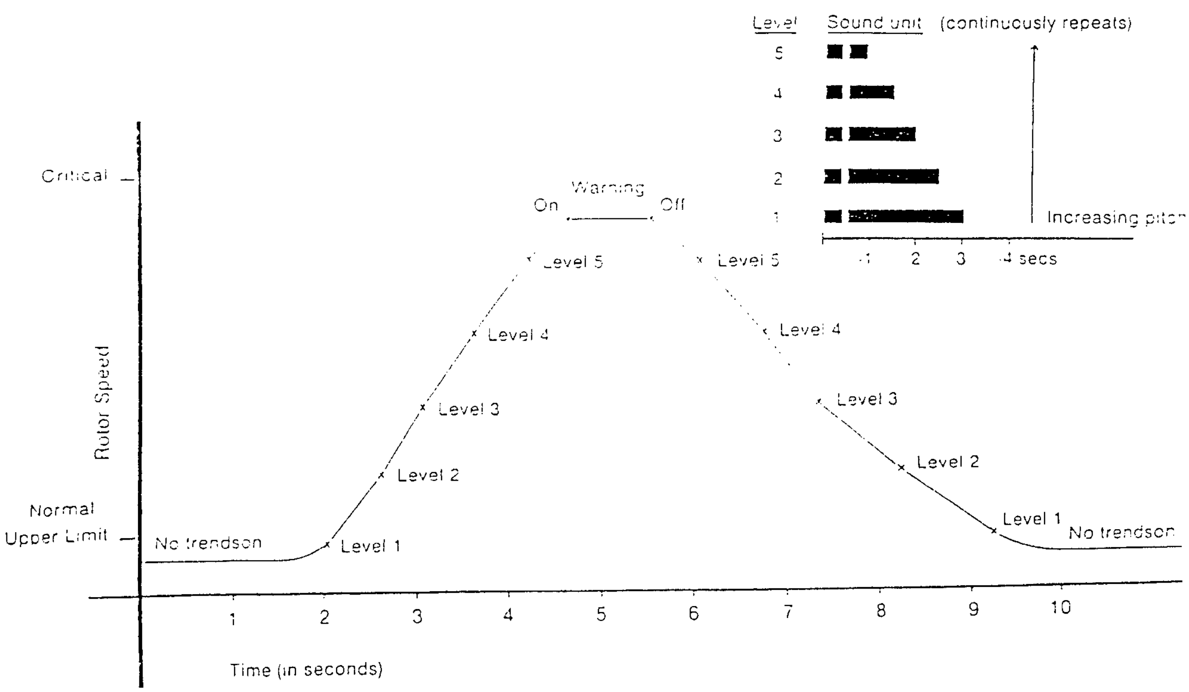


Figure 10 *Presentation Structure for a trend indicating sound.*

EXTENDING THE FREQUENCY RANGE OF EXISTING AUDITORY WARNINGS.

R.D. Patterson and A.J. Datta
MRC Applied Psychology Unit
15 Chaucer Road, Cambridge, CB2 2EF, U.K.

1. SUMMARY

This paper discusses two projects involving auditory warnings in military helicopters and fixed-wing aircraft. The first project reports methods to increase the frequency range of the existing DRA auditory warnings without changing their sound quality. The need to extend the frequency range arose from the requirement for "out-of-head" localisation of warning sounds. In the second project, the purpose was to develop a new class of sounds to be used as threat warnings. The aim was to make threat warnings that had a distinct sound quality as a set but which were, at the same time, separately identifiable.

2. INTRODUCTION

Several years ago, at the request of RAE FS9 (now DRA Ae FS9), the Applied Psychology Unit (APU) prepared a set of 12 auditory warning sounds for use in military helicopters to signal potential problems in flight systems. The sounds were prepared in accordance with guidelines that are now summarised in Patterson (1990), and they were tested to ensure a lack of confusability by Munger and Rood of DRA Ae FS9.

In the original flight-systems warnings most of the energy was below 4000 Hz. There was a need to increase the frequency range of the auditory warnings to 12000 Hz to make them localisable in the advanced audio display unit envisaged by DRA Ae. At the same time, it was important to change our perception of the sounds as little as possible since the existing warnings were already installed in operational aircraft.

There was no way of increasing the frequency range of the existing auditory warnings without producing some noticeable change in their sound quality. However, we noted a) that the temporal pattern of a complex sound is a major determinant of its character, and b) the main contribution of the high frequency components was to brighten the timbre of the sound. This suggested that a practical solution might be to add high-frequency energy with the same temporal envelope to each of the existing warning sounds. There appeared to be three ways of solving this problem:

Envelope Filling:
The envelope of the existing warning sound was extracted and applied to a set of high frequency harmonics. Then the two complex waves were combined in appropriate proportions.

Nyquist Whistling:
The existing warning sound was digitised at a rate just above that required by the main energy band. Then they were replayed without an anti-aliasing filter. This added high frequency energy in the form of a reflection of the original spectrum about the half sampling rate. This gave the sounds a distinctive whistling character. The sampling rate has to be tuned to get the right degree of whistling.

Fine Structure Doubling:
The waveform of the existing warning sound was segmented into cycles from one zero-crossing to the next and each cycle was replaced with two compressed versions of the cycle that fitted the same cycle time. Then the time compressed sound and the original sound were combined in appropriate proportions to produce the new warning.

This paper describes the algorithms and tools used to produce a large set of prototype warning sounds based on these generation techniques, and the listening procedures used to evaluate the prototypes in preparation for selecting a final set. The tools are available from the software package associated with the Auditory Image Model (AIM) (Patterson et al., 1995).

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

3. ENVELOPE FILLING

The envelope filling technique is a form of amplitude modulation in which one signal, the 'carrier', is multiplied by a second signal, the 'modulator'. The envelopes of the original auditory warnings are the modulators in this case and the carrier is a set of high frequency harmonics. The new warning is produced by multiplying the modulator by the carrier. To start with, analog recordings of the existing warnings were digitised at a sampling rate of 20000 points per second; the half sampling rate was well above the highest frequency in these warning sounds.

The first task in the *envelope filling* method, as the name suggests, was to extract the envelope from the original warning sound. The AIM routine for generating a spectrogram of a wave was used to extract the envelope. The filter-bank and compression were turned off; full wave rectification and low-pass filtering were turned on. The decay time of the low-pass filter was kept short (5 ms) to ensure that brief dips in the envelope of the original sound were preserved.

A C program, *phasesine.c*, was used to generate the high frequency harmonics. Twenty harmonics of 250 Hz, from 6000 Hz to 11000 Hz, with random phase, were added to produce a high frequency complex tone (referred to as **hfh**). The reason for adding in random phase was to avoid creating sound waves with large peak factors. Shell scripts were used for the generation of modified signals. The script file is *generate-highhar*, i.e "generate high harmonics".

The envelope and the high frequency harmonics were multiplied to form an amplitude modulated signal. The resultant wave was divided by the maximum value of hfh, to normalise it to the height of the envelope of the original sound. Finally, the resultant amplitude modulated waveform was added to the original warning to produce the new prototype warning. Three forms of each prototype warning were produced with the level of the added harmonics having the same, one half, or one quarter of the energy of the original warning.

4. NYQUIST WHISTLING

The Nyquist Whistling technique is a novel use of the Sampling Theorem which normally specifies the minimum sampling rate for adequate representation of a continuous signal. The critical sampling rate is twice the rate of the highest frequency component in the signal, and it is called the *Nyquist Frequency*. When sounds are recorded at too low a rate, or when they are replayed without an anti-aliasing filter, the original sound is accompanied by a whistling sound at high frequencies. The whistling can be explained by the spectrum of the recorded sound. When a signal is digitised, a copy of the spectrum in the region below the half-sampling frequency appears reflected in the spectrum between the half-sampling frequency and the Nyquist frequency. The effect is known as aliasing and it is an unavoidable byproduct of digitising a continuous wave. Normally, an anti-aliasing filter is used to remove the high frequency portion of the spectrum.

4.1 The Nyquist Whistling technique

The auditory warnings were recorded at sampling rates of 8000, 10000, 12000 and 16000 samples per sec. The aim was to locate the sampling rate which would just pass the energy of the auditory warning (without losing too much information), that is the effective Nyquist frequency. When the warning recorded at 16000 samples/sec was replayed, it sounded identical to the original, indicating that most of the energy in the auditory warnings lay below 8 kHz.

The Datlink interface used to record and play the sounds has a built-in anti-aliasing filter. To neutralise it, each point of the sampled warning was copied n times and played back at n times the recording rate, at which point, the whistling effect becomes audible. Each of the twelve auditory warnings, were passed through a routine called *ntimes*. The function *ntimes* took the auditory warning as input and the output was each point of the digitised warning written **n-times** in the output stream (where n was 1, 2, 3, ...). The modified warning was played back at **n-times** the recording rate. For example, warnings recorded at a sampling rate of 12 kHz, were passed to *ntimes* with argument 4, writing each point 4 times onto the output, and the modified warning was played back at a sampling rate of 48 kHz. This technique neutralised the anti-aliasing filter on the datlink interface and enabled one to hear the Nyquist whistling.

5. FINE STRUCTURE DOUBLING

Theoretically, this method is the most appropriate for adding high frequency components and producing minimal change in sound quality because it produces a sound like the octave of the original which should blend well with the original in terms of sound quality. However, most of the original warnings had highly irregular wave shapes and rapidly varying cycle times (periods). As a result, they strained the algorithm even with a 48000 Hz sampling rate, and the resulting distortion rendered some of the modified warnings unusable.

The aim of the technique was to replace each cycle of the digitised wave by two compressed copies of the cycle. The warning was digitised at a high sampling rate (48000 Hz) and the wave was divided into cycles from one zero crossing to the next. After a cycle had been isolated, every other point was dropped before doubling so that the total number of points per cycle remained constant. If there was a mismatch of one point at the end of the doubled cycle, the distortion was audible. For a cycle with even numbered points, dropping alternate points and doubling the compressed cycle was straight forward. For a cycle with an odd number of points, the last point of the compressed cycle was dropped when it was copied for the second time, to keep the cycle length constant. The original and double waveforms were divided by two so that when they were added, they stayed within the two byte limit.

6. LISTENING TESTS

Hhplay was the listening tool for the "envelope filling" method. The original warning was played first, followed by the high frequency *amplitude modulated* sound on its own. Then the original warning and three prototypes were played: 1) the original plus the hfh carrier, 2) the original warning plus the hfh carrier divided by 2, and 3) the original warning plus the hfh carrier divided by 4. The whole sequence could be repeated *n* times by specifying n as the second argument of the shell tool **hhplay**.

Nqplay was the script file for listening to the signals generated by the "Nyquist whistling" method. The original warning, recorded at 20000 samples/second was played first, followed by the versions recorded at 12000 samples/sec, 10000 samples/sec and 8000 samples/sec. Then the original warning and three prototypes were played: 1) the warning recorded at 16000 samples/sec with each point written three times, 2) the warning recorded at 12000 samples/sec with each point written four times, and 3) the warning recorded at 8000 samples/sec with each point written six times. Finally the original and the warning recorded at 10000 samples/sec with each point written twice were played. The number of repetitions could be specified by the listener as the second argument of the tool **nqplay**.

The final listening tool was **dcplay**. The original warning and three prototypes were played: 1) the original plus the cycle doubled signal, 2) the original plus the cycle doubled signal divided by two, and 3) the original plus the cycle doubled signal divided by four. The listener could specify the number of repeats using the second argument of **dcplay**.

7. LISTENING RESULTS

Listening tests were performed with the staff from APU and DRA. Judgements of the effectiveness of each method of extending the frequency range were made for each of the existing warnings, and the best value for the parameters was noted. The full set of judgements is recorded in Patterson and Datta (1994). This section describes a subset of the results - primarily the best new warnings but with comments on a few of the worst.

7.1 Auditory Warning Number 1

The result of envelope filling was not good for this warning. In the original sound, the envelope fluctuated rapidly and the spectrum changed rapidly with the envelope. When the envelope was filled with static high-frequency harmonics, the rapid fluctuations were less audible because there was no concomitant spectral change. The two signals sounded like two separate sources. There was something very shrill about the hfh component in this case. It drew the listener's attention well, but it quickly became aversive.

The Nyquist whistling method did a surprisingly good job of adding high-frequency energy while maintaining the good aspects of the original sound quality. In fact, Nyquist whistling worked quite well for most of the warnings. The only decision was to choose the best degree of whistling, that is, the best sampling rate for the extension in the blend. Perceptually, the blend with the 10 kHz extension seemed best.

The distortion introduced by fine structure doubling technique was particularly intrusive with this warning. The added components sounded like a totally different source that had only the modulator in common.

In summary, Nyquist whistling with the 10 kHz extension seemed best for warning 1.

7.2 Auditory Warning Numbers 2, 4, 5, 10

These warnings were similar to the first and produced a similar preference for Nyquist whistling with a 10 kHz cutoff.

7.3 Auditory Warning Number 3

This warning was well suited to the envelope filling method. The envelope of the original sound had well spaced fluctuations. Hence, when the high-frequency harmonics were added to the warning the effect was pleasant.

Nyquist whistling also worked well as did the fine structure doubling. The blend with extension 10 kHz sounded very good as did the blend with extension 8 kHz. Any one of the three methods would do in this case.

7.4 Auditory Warning Number 6

Envelope filling worked well for this warning with hfh at half the level of the primaries. Nyquist whistling was acceptable; the blend with 10 kHz extension seemed the best.

Fine structure doubling introduced distortion in the extension. When blended to produce new warnings, the added components produced a change in the timbre which was intrusive even when the relative level of the extension was low. The extension also changed the perceived urgency inappropriately.

The original warning had well spaced pulses and, as a result, all three methods worked in the sense of producing recognisable blends. Nevertheless, envelope filling and Nyquist whistling produced better warnings than cycle doubling.

7.5 Auditory Warning Number 7

Both the Nyquist whistling and the fine structure doubling produced a good result. The extensions were not perceived as separate sources in the blend. Rather, they brightened the timbre of the warning and gave it extra distinctiveness. Since the cycle doubling method did not produce good blends as often as Nyquist whistling, it would make sense to use cycle doubling in this case to increase the distinctiveness of the warnings within the set.

7.6 Auditory Warning Number 8

The original warning sounded like a calliope because it had a breathiness that was reminiscent of air whistling through pipes. The envelope filling extension had no breathiness whatsoever, and in the blends the extension reduced the overall breathiness considerably. As a result, the blends were quite different in character from the original.

The quality of the Nyquist whistling blend with the 10 kHz extension was probably better than the original. But, all the blends had a "chirp" that increased their distinctiveness and made all of them more acceptable than the original.

The cycle doubling method produced extensions with a strong breathiness, and so the prototypes with cycle doubled extensions had more breathiness than the original. The effect was highly satisfactory, and since the cycle doubling method did not produce good blends as often as Nyquist whistling, it makes sense to use it in this case as well.

7.7 Auditory Warning Number 9

This warning had well spaced pulses, yet the envelope filling method was not very successful for this sound. One could perceive the presence of two separate sources which made it quite different from the original. The shrill character of the extension made it stick out in the blend. New warnings produced by Nyquist whistling technique had a chirping effect which threatened to dominate the character of the warning.

Cycle doubling produced extensions that were not heard as separate sources and which seemed to intensify the natural character of this warning sound. The extensions improved the sharpness of the warning while preserving the fundamental character of the sound. Fine structure doubling produced the best overall results and the different blends all seemed equally acceptable.

7.8 Auditory Warning Number 11

The blends produced by envelope filling were all quite bad. Two separate sources were heard because the extension did not have the strong frequency sweep of the original. The timbre of the extension stuck out and made the prototype sound very different from the original. The Nyquist whistling method actually reduced the level of this warning sound. This suggested that the original sound already had high frequency energy, and that this energy was removed in the recording process when the lower cutoffs were used. If this was the case, there was no need to modify this warning sound.

Fine structure doubling produced the best outcome here. The high-frequency extension produced minimal disruption of sound quality in the blends. The extension increased the sharpness of the sound with the introduction of some noisiness which was generally acceptable.

7.9 Auditory Warning Number 12

The envelope filling technique did not work for this warning sound since the original sound had no amplitude modulation. The original sound drew attention through frequency modulation, which could not be captured by this procedure. Hence the extension was clearly distinguishable as a separate source in all three blends.

The Nyquist whistling effect was barely audible in blends of this warning sound, indicating that the upper frequencies in this warning lie well below the lowest half-sampling rate (8 kHz) and the reflected part of the spectrum is limited to a small region around the Nyquist frequency. In any event, it did not produce a good warning sound.

Fine structure doubling produced the best outcome with this warning; indeed, it was probably the best result for this method with any of the 12 warnings. When the high-frequency energy was added, it barely changed the timbre of the warning. The brightness was increased and that was about the only audible change. The technique worked well at all the three levels in the blends.

8. AUDITORY WARNINGS WITH TEMPORAL ASYMMETRY

In this phase of the project, the purpose was to develop a new class of sounds to be used as threat warnings. There was also the constraint that they needed to be integrated with the "aircraft systems" warnings and "flight parameter" warnings to form a well structured warning suite. This led to the investigation of new class of sounds with temporal asymmetry in the envelope. These warnings have a timbre or sound quality, different from the existing warnings, so they are distinguishable as a class. Instead of synthesis in the frequency domain, the new sounds are generated in the time domain. Varying individual parameters produces a range of similar sounds with varying degrees of urgency. Unique combinations of envelopes and carriers can be used to produce identifiable, dedicated attensons.

Temporally asymmetric envelopes were applied to a carrier to produce a distinctive new sound. A degree of jitter was also added to the envelope period, or one of the other parameters, to increased distinctiveness and urgency. We report the results of searching the space of parameters and values to find appropriate sounds for aircraft threat warnings. Various complex carriers were used which broadly fell into two categories; a) complex tones, where harmonics or octaves with random phases were used as constituent sounds, and b) iterated ripple noise (**IRN**) which is constructed from a random noise by delaying a copy of the noise, adding it to the original, and iterating or repeating the process a number of times.

Various envelope shapes were considered while investigating asymmetry. Damped and ramped envelopes were chosen as the staring point, since the effect of shape was not dramatic and the damped and ramped envelopes have been studied systematically by Patterson (1994a, 1994b) and Irino and Patterson (1996). The variables which affect damped/ramped sounds are the envelope period, the half life of the exponential decay, and the amplitude and the floor level where the damped/ramped envelope ends. Randomness (or jitter) was introduced to each of these parameters to enhance distinctiveness. The search space was vast and so we began by developing tools to explore the space systematically, taking one parameter and one carrier at a time, to find a reasonable range.

9. A STRUCTURED LISTENING TOUR THROUGH THE SPACE OF ASYMMETRIC SOUNDS

The time-asymmetric envelopes form the basis of the new class of warnings. Four carriers were used to produce four distinct subclasses of sounds. Tools were prepared to present large numbers of these sounds to listeners in a convenient form. The parameters were varied systematically and perceptual descriptions of the sound qualities were noted. The range was surprisingly large so a subset of the tables of results are presented below. The sounds marked with asterisks show the ones which could form the basis for attensons.

9.1 Octave-spaced Harmonic Carrier *Damped Sounds*

Half-life	Period	Timbre
4ms	45ms	Drumlike organ clicks
4ms	90ms	Pizzicato organ notes
4ms	180ms	Stronger yet brief organ notes
8ms	45ms	Organ component stronger and rapid flutter
**8ms	90ms	Organ component stronger and flutter
**8ms	180ms	Organ component stronger with slower flutter
16ms	45ms	Tone begins to dominate
**16ms	90ms	Metallic organ taps
16ms	180ms	Metallic organ taps with more damping heard
32ms	45ms	Too tonal, weak clicks
**32ms	90ms	Metallic bell sounds
32ms	180ms	Bell sounds

9.2 IRN (lag 16ms) Carrier *Damped Sounds*

Half-life	Period	Timbre
4ms	45ms	Rapid brief clicks
**4ms	90ms	Brief clicks
4ms	180ms	Brief clicks
8ms	45ms	Snare drum effect
**8ms	90ms	Snare drum effect
8ms	180ms	Snare drum effect
**16ms	45ms	Propeller Plane
16ms	90ms	Propeller plane with slow rotation
16ms	180ms	Dull sound
**32ms	45ms	Loud propeller plane
32ms	90ms	Cylinder helicopter
**32ms	180ms	Slow Piston like effect

9.3 IRN (lag 16ms) Carrier *Ramped Sounds*

Half-life	Period	Timbre
4ms	45ms	Noisy flutters/clicks
4ms	90ms	Noisy flutters/clicks
4ms	180ms	Noisy clicks
8ms	45ms	Noisy clicks with little tone
8ms	90ms	Noisy clicks with little tone
8ms	180ms	Noisy clicks with little tone
16ms	45ms	Ships funnel
**16ms	90ms	Cylinder helicopter
16ms	180ms	Piston with stronger tone
**32ms	45ms	Loud ships funnel
32ms	90ms	Piston effect pronounced
32ms	180ms	Strong piston with tone

10. THE EFFECT OF JITTER

The effect of jitter seemed to be largely orthogonal to sound quality for these sounds. When jitter was gradually introduced to one of the envelope parameters, the perception of the sound did not change suddenly to that of a new source. As a result, jitter was omitted in the initial, parametric, listening tests. Then, once an interesting sound was identified, randomness was introduced in one of the parameters to accentuate the distinctiveness of the sound.

We started our search by creating sounds having a wide range of jitter in exponential steps. The initial range was from 1% to 90% with the steps being 1%, 10%, 30%, 50% and 90%. In general:

a) The range 1%-10% hardly produced any noticeable difference in the sounds.

b) Jitter above 80% produced sounds with crackling distortions which were generally disruptive.

c) Jitter in the range of 10%-30% produced sounds where the effect was gradually noticed. Though psychophysically important, we thought that this effect was not strong enough to produce distinctiveness in attensons.

d) The ideal range seemed to be 30%-70%. So, exploration of the space was focused on sounds with 30%, 50% and 70% jitter.

We listened to a wide range of the asymmetric sounds with jitter, separately, in all four parameters. It was noted that those sounds which were dull to start with (no asterisks) did not become good attensons by virtue of the addition of jitter. However, for those that had already been chosen as potential attensons, jitter tended to enhance their attention gaining quality. The effects of jitter in the four parameters are summarised below:

Half-life

Random fluctuations in the half-life seemed to impart rhythmic patterns to the sounds. As the percentage of variation increased, the rhythm became quite strong but it was never disruptive. For example, the octave-spaced harmonic carrier with a damped envelope (16ms half-life), sounded like metallic organ taps. With a jitter of 50%, it was metallic organ taps with a rhythmic pattern – an effect which would probably enhanced retention of the sound quality in memory.

Envelope Period

Variations in the envelope period added jumpiness or hesitation to the sound quality. Listening tests did not produce sounds which seemed better than the non-random sounds or half-life jittered sounds. At higher values, crackling distortion was added to the sounds.

Amplitude

As expected, overall amplitude variation randomly changed the loudness of the individual pulses making up the sounds. It was an interesting effect, but probably not useful for attensons.

Floor

Floor variation produced sounds which seemed to have effects of both envelope period jitter and amplitude jitter. There was some hesitation with random loudness variations. Two levels of floor variations of 60% and 90% were studied. At the higher end, crackling distortion was present. At the 60% level, the irregularity did not create perceptions different from the ones discussed above.

Following these observations, we chose to concentrate on the half-life variations for the current attensons.

11. THE GENERATION OF NEW SET OF THREAT WARNINGS

Research by the DRA and APU has shown that increasing the number of auditory warnings beyond six or seven becomes counter-productive (Patterson, 1982). Moreover, warnings which have simple temporal patterns are confused (Patterson, 1990). Anticipating the need for threat warnings, the DRA has produced a four level structure for warning sounds to ensure correct coding of urgency. Level I has the lowest urgency and level IV the maximum urgency. The levels for radar threats are:

Level I Undetected Search (Radar)

Level II Acquisition by Radar

Level III Tracked by Radar

Level IV Missile or Gun Launched

The structure was implemented, as specified by DRA, as follows:

Level I advisory attenson + voice message

Level II dedicated attenson + "Missile."

Level III dedicated attenson + "Unknown."

Level IV dedicated attenson + "Laser!"

For level I, level II and level III threats, we implemented a radar like sweeping sound. Back-to-back damped and ramped envelopes produce a sound with asymmetry that can be controlled in a useful way, but the sharp peak where the two components of the envelope meet produce a sharp click in the sound. This suggested that the most appropriate envelope shape would be the *roex*. The idea of the roex envelope came from work in the spectral domain on auditory filter shapes (Patterson, 1976). The *ro*unded *ex*ponential (roex) shape has a rising exponential onset, a rounded top and a decaying exponential offset. The rounded top was introduced to make the filter flat at its centre frequency. When translated to the time domain the rounded top prevents the unpleasant click just as it makes the derivative of the filter function smooth at its centre frequency. The ramped half-life and the

damped half-life of the roex are independent parameters. Hence we could produce smooth sounding, time asymmetric envelopes - the central theme in these new sounds. To make a low-urgency sound, a roex envelope with long half-lives was applied to a low-pitched carrier. Increasing the carrier frequency, decreasing the roex half-lives, and repeating the pulses rapidly generated warnings with greater urgency. So with the same type of attenson, changing one or two parameters changes the urgency (Patterson, 1990).

A prototype set of threat-warnings was generated and recorded. The carrier used with the roex envelopes was *I*terated *R*ipple *N*oise (IRN).

Level I, "Undetected search", was made with a roex envelope with period 800 ms, ramped half-life of 64 ms and damped half-life of 128 ms, with 6 seconds of silence between the pulses (in total, 4 pulses). The carrier was an IRN with 16 iterations and a 16-ms delay (low pitch).

Level II, "Acquisition", was made with a roex envelope with period 400 ms, ramped half-life of 32 ms and damped half-life of 64 ms, with 3 seconds of silence between the pulses (in total, 4 pulses). The carrier was an IRN with 16 iterations and an 8-ms delay.

Level III, "Acquisition", was made with a roex envelope with period 200 ms, ramped half-life of 16 ms and damped half-life of 32 ms, with no silence between the pulses (in total, 4 pulses). The carrier was an IRN with 16 iterations and a 4-ms delay.

Two level IV sounds were made. One for the "Missile" message and the other for the "Gun" message.

Level IV, "Urgent missile", was made as a two part warning. The first part was a dedicated attenson constructed with a ramped envelope with period 90ms, 30% jitter in pulse time and a carrier of IRN with 16 iterations and a 2-ms delay (high pitch). The second part which identified the missile, was made of a long period ramped sound (180 ms) and an IRN carrier having an 8-ms delay.

Level IV, "Urgent gun", was also made as a two part warning. The first part was a dedicated attenson constructed with a ramped envelope with period 90ms, 30% jitter in pulse time and a carrier of IRN with 16 iterations and a 2-ms delay (high pitch). The second part, identifying a gun, was made with a long period damped sound (180 ms) and a random-phased harmonic carrier.

12. CONCLUSION

The set of existing DRA warnings with extended frequency range, and the new set of threat warnings are currently under evaluation by DRA.

13. REFERENCES.

Irino, T. and Patterson, R.D. (1996). "Temporal asymmetry in the auditory system," J. Acoust. Soc. Am. 99, 2316-2331

Patterson, R.D. (1976) "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am., 59, 640-654

Patterson, R.D. (1982) "Guidelines for auditory warning systems on Civil Aircraft," Civil Aviation Authority, Paper 82017, London

Patterson, R.D. (1990). "Auditory warning sounds in the work environment," Phil. Trans. R. Soc. Lond. B 327, 485-492.

Patterson, R.D. (1994a). "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am. 96, 1409-1418.

Patterson, R.D. (1994b). "The sound of a sinusoid: Time-interval models," J. Acoust. Soc. Am. 96, 1419-1428.

Patterson, R.D., Allerhand, M., and Giguere, C., (1995). "Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. 98, 1890-1894.

Patterson, R.D. and Datta, A.J. (1994). "Extending the frequency range of existing warning sounds," AAM HAP Auditory Warnings: Progress Report No. 1, APU Contract Report.

THE ABILITIES OF LISTENERS TO LOCALISE DEFENCE RESEARCH AGENCY AUDITORY WARNINGS

Russell L. Martin
Simon P.A. Parker
Ken I. McAnally
Simon R. Oldfield
Air Operations Division
Aeronautical and Maritime Research Laboratory
506 Lorimer St.
Fishermens Bend, 3207, Australia

SUMMARY

The Defence Research Agency (DRA) in collaboration with the Applied Psychology Unit at Cambridge University and the Institute of Sound and Vibration Research at Southampton University have designed a set of 12 auditory warnings for use in military aircraft. These warnings have recently been modified to extend their high-frequency content in an attempt to increase the accuracy with which they can be localised and therefore enhance their suitability for use in conjunction with a 3D audio display. We have evaluated the abilities of listeners to localise a sample of the original DRA auditory warnings and their high-frequency versions. Eight subjects localised broadband noise and five original warnings when presented from a loudspeaker at each of 40 locations ranging from -40 to +40° azimuth and -50 to +50° elevation in their frontal hemifields. Subjects were divided into two groups of four subjects each on an age basis: 22-28 year olds and 33-48 year olds. Consistent with the results of previous studies, an average localisation error of about 5° was observed for a train of three 150 ms bursts of broadband noise for both age groups. Average localisation errors for the five original auditory warnings, however, were much larger and varied from about 10 to 25° for the younger subjects and 15 to 30° for the older. Four subjects, aged from 23-39 years, then localised three modified versions of two of these original warnings. Of the three modification methods employed, only one (fine-structure doubling) produced stimuli that were localised more accurately than their original versions. The improvement in localisation accuracy for stimuli modified by this method resulted primarily from an improvement in the accuracy with which the elevation of the stimulus could be determined.

1. INTRODUCTION

About ten years ago the Defence Research Agency (DRA) in collaboration with the Applied Psychology Unit at Cambridge and the Institute of Sound and Vibration research at Southampton University designed a set of 12 auditory warnings, known as attensons, for use in rotary and fixed-wing military aircraft [1]. These warnings were developed in keeping with design principles described in a previous article in this volume [2] and evaluated to ensure they could be reliably discriminated from one another. While originally designed as flight-systems warnings, these sounds may eventually be used in conjunction with a three-dimensional (3D) audio display for the purpose of conveying spatial information to aircrew. As such, the accuracy with which they can be localised needs to be considered.

Previous research has established that auditory signals must be broadband and contain energy above about 4 kHz for them to be accurately localised when presented from a loudspeaker in a free-field [e.g., 3, 4, 5, 6]. As most of the DRA attensons referred to above incorporate little energy above this frequency, it could be expected that they would be difficult to localise and therefore unsuitable for use with a 3D audio display. In view of this, these original warnings have been modified, using three different methods, to increase their high-frequency content [7]. As part of a collaborative project carried out under the auspices of the Technical Co-operation Programme Sub-group U Technical Panel 7 we have evaluated the abilities of listeners to localise a sample of the original DRA attensons and their high-frequency versions. Our study was divided into two parts: in the first we tested the abilities of listeners to localise five of the 12 original attensons and in the second we examined their abilities to localise the three modified versions of two attensons. Preliminary observations suggested that the accuracy with which listeners could localise the original attensons depended on their age, presumably as a consequence of high-frequency hearing loss associated with the aging process. Accordingly, we decided to evaluate the accuracy with which the original attensons could be localised using subjects of two different age groups.

2. METHOD

2.1 Subjects

Eight subjects participated in this study. They were divided into two groups of four subjects each on an age basis. Subjects in the first group were aged from 22 to 28 years while those in the second were aged from 33 to 48 years. Hearing thresholds for 1, 4, 8, 12 and 16 kHz pure tones were determined for each subject and found to be within two standard deviations of age-relevant norms [8 (1-8 kHz), 9 (12 & 16 kHz)]. All eight subjects provided data for the first experiment in this study, which was concerned with localisation of the original DRA attensons. For the second experiment, concerned with localisation of the modified attensons, data was collected from four subjects only: one from the

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

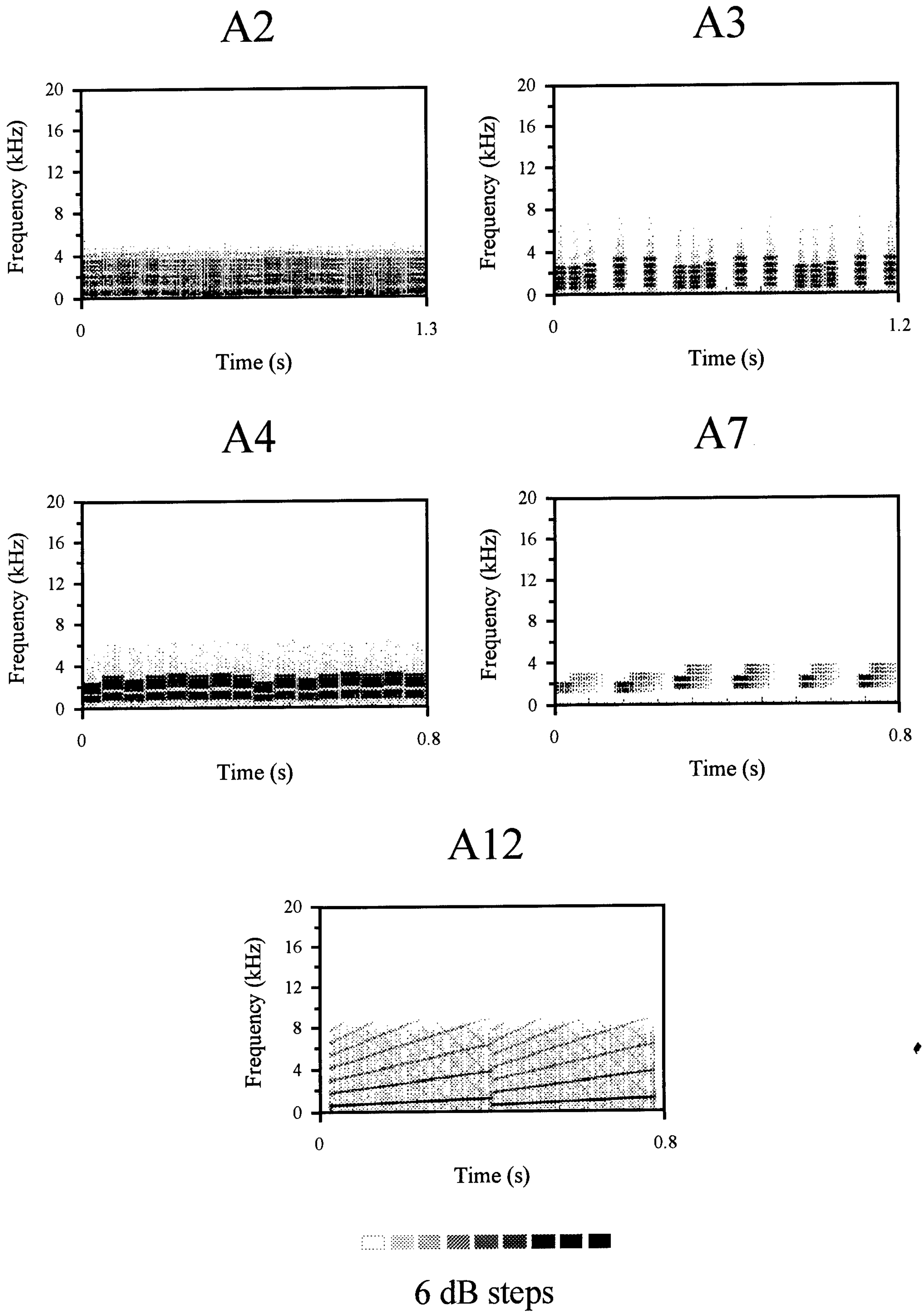


Fig. 1. Spectrograms of five original attensons. A2, attenson 2; A3, attenson 3; A4, attenson 4; A7, attenson 7; A12, attenson 12.

22-28 year old group and three from the 33-48 year old group.

2.2 Stimuli

The five original DRA attensons tested in this study were selected in view of the need to evaluate a subset of the full set of 12 that was representative with respect to spectral and temporal properties. Spectrograms of the five original attensons chosen for testing are presented in Figure 1. Attenson 2 (A2) comprised of an almost continuous burst of energy of frequencies ranging from 0 to about 5 kHz and was chosen because of its average spectral content (with respect to the spectral content of all 12 original attensons) and continuous nature in the temporal domain. Attenson 3 (A3), on the other hand, was chosen as an example of an attenson consisting of a series of distinct pulses separated by periods of silence and contained energy up to about 6.5 kHz. Attenson 4 (A4) also contained energy up to about 6.5 kHz but was almost continuous in the time domain. Attenson 7 (A7) contained no energy above about 4 kHz and was chosen for testing because of its particularly limited high-frequency content. The last original attenson chosen for testing was attenson 12 (A12), which consisted of a series of continuous frequency sweeps. Of all original attensons tested it had the most energy at high frequencies, extending up to about 8.5 kHz.

The original DRA attensons were recently modified to extend their high-frequency content using three different methods: envelope filling, Nyquist whistling and fine-structure doubling. These methods have been described in detail in a previous article in this volume [7]. As each method appeared to increase the high-frequency content of the original attensons in a distinctive but characteristic fashion, it was decided to test the accuracy with which the modified versions of only two of the original attensons could be localised. Spectrograms of the modified versions of these two attensons, attensons 2 and 7, are shown in Figure 2. Modification method 1, envelope filling, added three narrow bands of energy centred at about 4, 6 and 8 kHz to both of the original attensons (A2M1 & A7M1). Method 2, Nyquist whistling, was the most successful in terms of increasing high-frequency content and added four broad bands of high-frequency energy to both attensons, with the highest band containing components up to 17-18 kHz (A2M2 & A7M2). Method 3, fine-structure doubling, added components to attenson 2 at frequencies up to about 12 kHz and to attenson 7 up to about 16 kHz (A2M3 & A7M3). In the case of this method, however, the high-frequency energy added to the original attensons was distributed more-or-less evenly up to the new high-frequency cut-off. That is, modification method 3, unlike methods 1 and 2, produced a signal that had no regions in its frequency profile where energy was absent.

A baseline measure of each subject's localisation ability was provided by measuring the accuracy with which they could localise broadband (0-20 kHz) noise. Each noise stimulus consisted of a train of three 150 ms bursts separated by 100 ms gaps. This stimulus was generated by an array processor card (AP2, Tucker-Davis-Technology) mounted in a host PC and converted to an analogue signal at a rate of 50 kHz by a 16-bit A/D converter (PD1, Tucker-Davis-Technology). Digital recordings of each attenson, which were stored on the host PC's hard disk, were converted to analogue signals at a rate of 44.1 kHz and all signals were amplified and presented to subjects at 62.5 dB sound pressure level via a Bose transducer (satellite speaker from the "acoustimass" system).

2.3 Procedure

The accuracy with which subjects could localise the original and modified attensons was evaluated using a stimulus presentation system that consisted of a loudspeaker mounted on a vertically-oriented, 1-m radius, semicircular boom. Subjects were seated such that their head was positioned at the centre of the sphere described by rotation of this boom about its vertical axis. Tests were conducted in dim light and the subject's view of the loudspeaker was obscured by an acoustically transparent, 0.95-m radius, cloth hemisphere supported by thin fibreglass rods. Each subject was required to localise each stimulus when presented from each of 40 locations ranging from -40 to +40° azimuth and -50 to +50° elevation in their frontal hemifield. Each stimulus was tested in a separate block of 40 trials that required about 20 minutes to complete. Subjects indicated where they perceived stimuli to emanate from by directing the light from a laser pointer mounted on a headband they were wearing at the relevant point on the surface of the cloth hemisphere. The position at which they pointed was determined by tracking the location and orientation of the laser pointer's tip using a six-degrees-of-freedom magnetic tracker (3Space Fastrak, Polhemus). The angle subtended by the two vectors pointing from the centre of the boom system to the real and perceived positions of the stimulus source provided a measure of localisation accuracy.

3. RESULTS

Localisation errors were averaged separately across the four 22-28 year old and four 33-48 year old subjects for broadband noise and each of the five original attensons and are shown in Figure 3. It can be seen that the pattern of localisation error across stimuli was similar for both age groups. Both groups localised the broadband noise stimulus most accurately and displayed a localisation error of a little over 5°, which is consistent with the results of previous studies [e.g., 10, 11]. The least accurately localised stimulus for both groups was attenson 7, for which the localisation error was about 25° for the younger group and about 30° for the older. Of the five attensons, attenson 12 was localised most accurately by both groups, with the younger group localising this stimulus almost as accurately as they localised broadband noise. Within both age groups, attensons 2, 3 and 4 were localised with about equal accuracy. Looking across age groups it can be seen that for all except the broadband noise stimulus the older group had markedly larger localisation errors. Statistical analysis of these data using an analysis of variance technique revealed significant main effects for both stimulus (Wilks' Lambda(5,2) = 0.003, $p < 0.05$) and age ($F(1,6) = 15.71$, $p < 0.05$). Subsequent planned comparisons indicated that each of attensons 2, 3, 4 and 7 was localised significantly less accurately than broadband noise by both groups of subjects (A2 young,

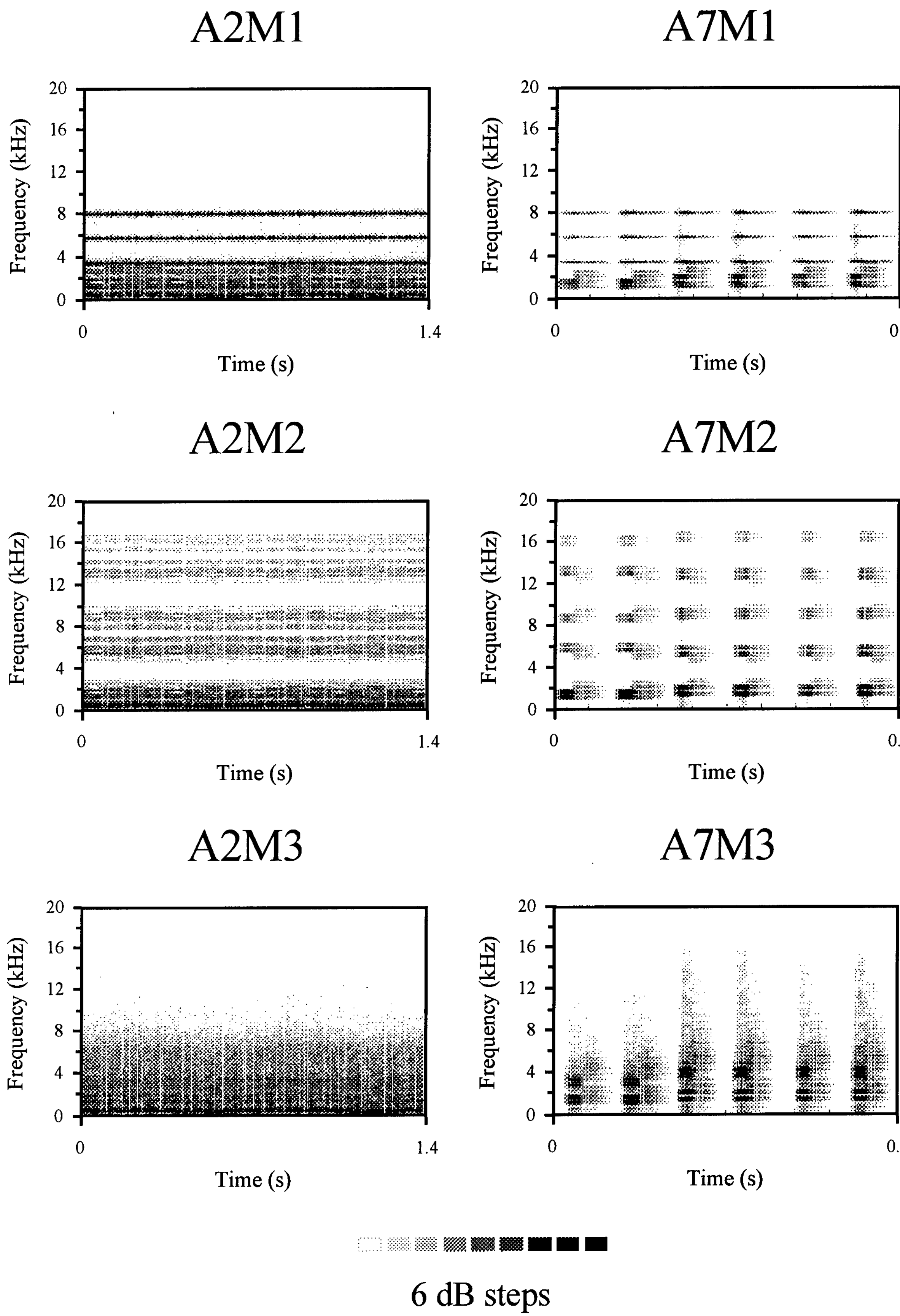


Fig. 2. Spectrograms of modified attensons. A2M1-3, attenson 2 modified by method 1-3; A7M1-3, attenson 7 modified by method 1-3.

F(1,6) = 16.38, $p < 0.05$; A2 old, F(1,6) = 84.29, $p < 0.05$; A3 young, F(1,6) = 11.02, $p < 0.05$; A3 old, F(1,6) = 29.13, $p < 0.05$; A4 young, F(1,6) = 27.19, $p < 0.05$; A4 old, F(1,6) = 116.33, $p < 0.05$; A7 young, F(1,6) = 110.57, $p < 0.05$; A7 old, F(1,6) = 181.58, $p < 0.05$) and attenson 12 was localised significantly less accurately than broadband noise by the older group only (F(1,6) = 18.98, $p < 0.05$).

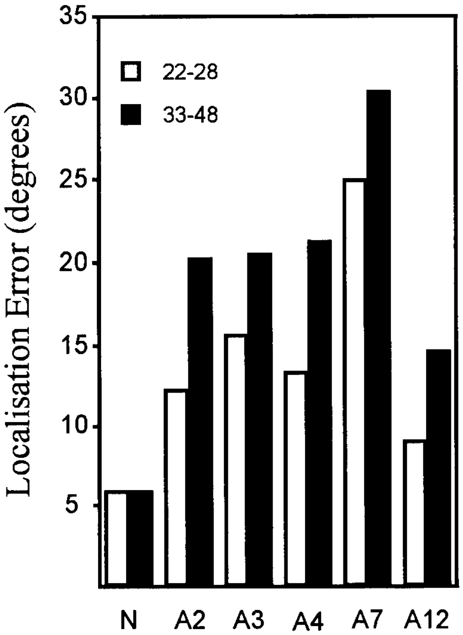


Fig. 3. Average localisation errors of subjects in the 22-28 and 33-48 year old groups for five original attensons. 22-28, 22-28 year olds; 33-48, 33-48 year olds; N, broadband noise; A2, attenson 2; A3, attenson 3; A4, attenson 4; A7, attenson 7; A12, attenson 12.

The azimuth and elevation components of these localisation errors are depicted in Figures 4 and 5, respectively. It can be seen that the azimuth errors (Fig. 4) were in all cases much smaller than the localisation errors, being mostly less than 5°, and showed little variation as a function of either age or stimulus. Statistical analysis confirmed that the effects of these two factors on azimuth error were insignificant. The corresponding elevation errors (Fig. 5), on the other hand, varied as a function of age and stimulus in a manner that greatly resembled the way the localisation errors varied. Elevation errors were generally larger for the older group, and for both age groups were largest for attenson 7 and smallest for attenson 12 and broadband noise. A statistical analysis of these errors revealed exactly the same pattern of effects as the one described above for the localisation errors (main effect stimulus, Wilks' Lambda(5,2) = 0.007, $p < 0.05$; main effect age, F(1,6) = 14.71, $p < 0.05$; A2 young, F(1,6) = 18.7, $p < 0.05$; A2 old, F(1,6) = 84.7, $p < 0.05$; A3 young, F(1,6) = 12, $p < 0.05$; A3 old, F(1,6) = 29.09, $p < 0.05$; A4 young, F(1,6) = 30.92, $p < 0.05$; A4 old, F(1,6) = 129.57, $p < 0.05$; A7 young, F(1,6) = 85.44, $p < 0.05$; A7 old, F(1,6) = 148.96, $p < 0.05$; A12 old, F(1,6) = 17.12, $p < 0.05$).

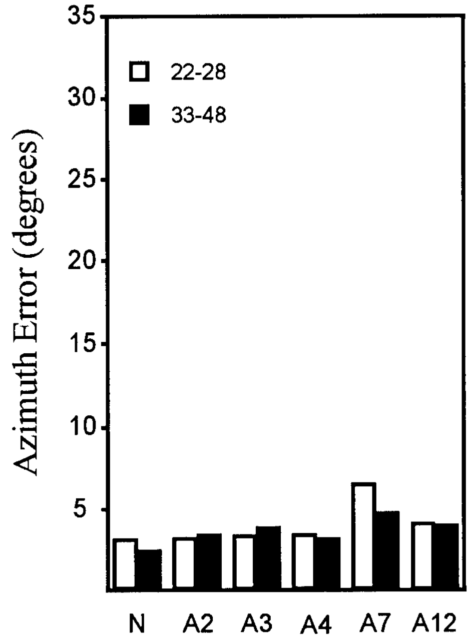


Fig. 4. Azimuth errors of subjects in the 22-28 and 33-48 year old groups for five original attensons. 22-28, 22-28 year olds; 33-48, 33-48 year olds; N, broadband noise; A2, attenson 2; A3, attenson 3; A4, attenson 4; A7, attenson 7; A12, attenson 12.

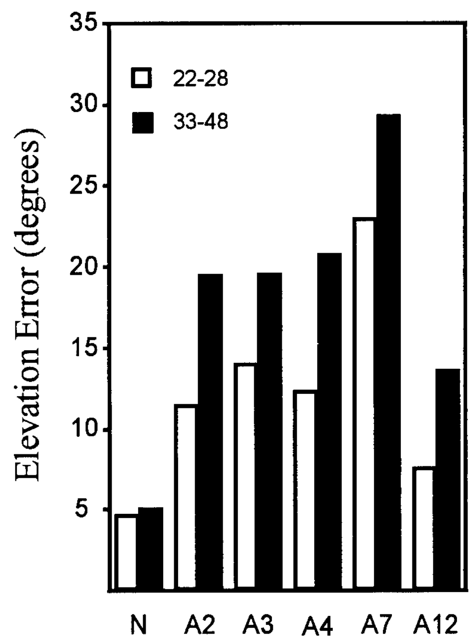


Fig. 5. Elevation errors of subjects in the 22-28 and 33-48 year old groups for five original attensons. 22-28, 22-28 year olds; 33-48, 33-48 year olds; N, broadband noise; A2, attenson 2; A3, attenson 3; A4, attenson 4; A7, attenson 7; A12, attenson 12.

Average localisation errors for the original and three modified versions of both attensons 2 and 7 are shown in Figure 6. It can be seen that the pattern of localisation error across modification method was identical for the two attensons. In both cases method 1 produced the least accurately localised modified version and method 3 produced the most accurately localised. The localisation errors for the versions produced by method 3 were of the order of 10°. A statistical analysis using an analysis of variance technique revealed significant main effects for stimulus ($F(1,3) = 62.67$, $p < 0.05$), reflecting the fact that errors were larger for attenson 7, and method (Wilks' Lambda(3,1) = 0.000, $p < 0.05$). A more detailed analysis in which each attenson was considered separately showed that modification method 3 was the only one to produce stimuli that were localised significantly more accurately than the original versions (A2, $F(1,3) = 18.46$, $p < 0.05$; A7, $F(1,6) = 85.32$, $p < 0.05$).

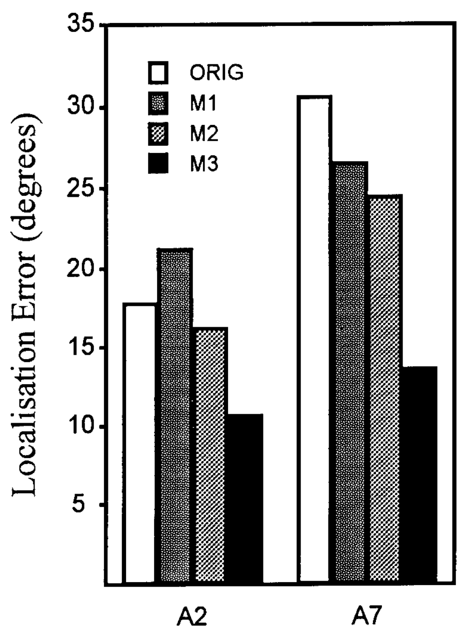


Fig. 6. Average localisation errors for modified versions of attensons 2 (A2) and 7 (A7). ORIG, original version; M1-3, versions modified using methods 1-3.

The azimuth and elevation errors associated with these localisation errors are shown in Figures 7 and 8, respectively. The azimuth errors (Fig. 7) were again much smaller than the localisation errors and did not vary significantly as a function of method. The corresponding elevation errors (Fig. 8), in contrast, were similar in magnitude to the localisation errors and varied as a function of stimulus in an identical manner. A statistical analysis of these elevation errors revealed exactly the same pattern of effects as the one described above for the localisation errors (main effect stimulus, $F(1,3) = 36.05$, $p < 0.05$; main effect method, Wilks' Lambda(3,1) = 0.000, $p < 0.05$; A2, $F(1,3) = 17.35$, $p < 0.05$; A7, $F(1,3) = 89.43$, $p < 0.05$).

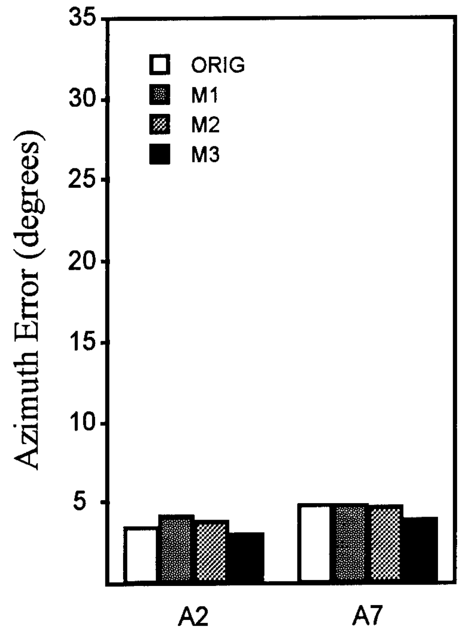


Fig. 7. Azimuth errors for modified versions of attensons 2 (A2) and 7 (A7). ORIG, original version; M1-3, versions modified using methods 1-3.

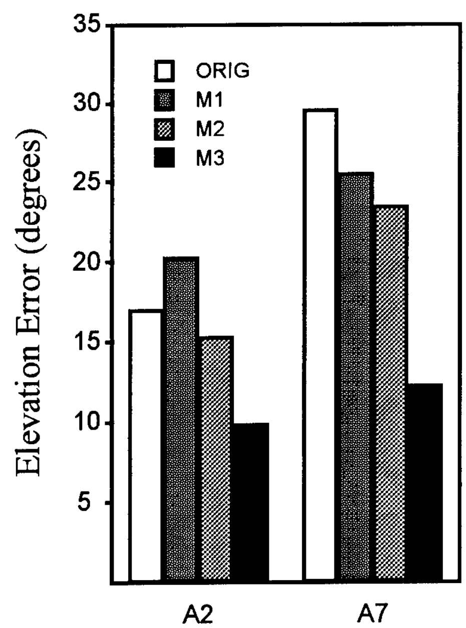


Fig. 8. Elevation errors for modified versions of attensons 2 (A2) and 7 (A7). ORIG, original version; M1-3, versions modified using methods 1-3.

4. DISCUSSION

The results of this study confirmed our expectation that the original DRA attensons would be difficult to localise

and indicate that, with the possible exception of attenson 12, the original attensons tested here are probably unsuitable for conveying spatial information via a 3D audio display. Whether any particular stimulus is regarded as suitable for this purpose, however, will of course depend on the degree of spatial resolution required when the stimulus is presented. The fact that the localisation errors associated with these original attensons were comprised mostly of elevation errors suggests that these stimuli were poorly localised because they lack high-frequency components. It is well known that the cues we use to determine the elevation of a source of sound result from the interaction of the head and pinnae with *high*-frequency soundwaves [e.g., 12].

Our expectation that the accuracy with which listeners would localise the original attensons would depend on their age was also confirmed, with the 22-28 year old subjects localising these attensons more accurately than the 33-48 year olds. This is not a surprising result and presumably reflects the lesser high-frequency hearing sensitivity of the older subjects. It does, however, demonstrate a need to evaluate the accuracy with which an auditory warning can be localised using subjects who have the same auditory sensitivity as those for whom the warning is being designed.

Two additional points are worth making on the basis of the results from the first part of this study. The first is that the accuracy with which each of the original attensons could be localised was predictable from its frequency content. Attenson 7 was the least accurately localised stimulus and the one chosen because of its particularly limited high-frequency content. Attenson 12, on the other hand, was the most accurately localised stimulus and the one that had the most energy at high frequencies. The second point is that the temporal structure of these attensons, that is whether they were continuous or pulsed, did not appear to affect the accuracy with which they could be localised.

The second part of this study showed that the accuracy with which the original DRA attensons can be localised can be improved using modification method 3 (fine-structure doubling) and that this improvement comes about because elevation errors are reduced. Furthermore, it showed that this method can improve the localisation of a very poorly localised stimulus, such as attenson 7, to the extent that it may be reasonable to use it in conjunction with a 3D audio display. From a more theoretical perspective, it is of interest that the localisation improvement associated with these modification methods does not appear to be related in a simple way to the extent to which they add high-frequency components. The modification method that increased high-frequency content the most was method 2, which failed to produce a stimulus that was localised more accurately than the original. It is possible that the way in which high-frequency content is distributed across frequency is critical in determining whether or not it will lead to improved localisation.

REFERENCES

1. Lower, M., Wheeler, P., Patterson, R., Edworthy, J., Rood, G., and Chillery, J., "The design and production of auditory warnings for helicopters. Vol. 1. Sea King", ISVR report AC527A, October 1986.
2. James, S.H., "Audio warnings for military aircraft", *in* "Audio Effectiveness in Aviation", AGARD CP, October 1996, Paper 7.
3. Butler, R.A. and Planert, N., "The influence of stimulus bandwidth on localization of sound in space", J.Acoust.Soc.Am., *19*, 1976, 103-108.
4. Hebrank, J. and Wright, D., "Spectral cues used in the localization of sound sources on the median plane", J.Acoust.Soc.Am., *56*, 1974, 1829-1834.
5. Rofler, S.K. and Butler, R.A., "Factors that influence the localization of sound in the vertical plane", J.Acoust.Soc.Am., *43*, 1968, 1255-1259.
6. King, R.B. and Oldfield, S.R., "The impact of signal bandwidth on auditory localization: Implications for the design of three dimensional audio displays", Human Factors, 1996, In press.
7. Patterson, R.D. and Datta, A.J., "Extending the frequency range of existing auditory warnings", *in* "Audio Effectiveness in Aviation", AGARD CP, October 1996, Paper 8.
8. Corso, J.F., "Age and sex differences in pure-tone thresholds", J.Otolaryngol., *77*, 1963, 385-405.
9. Stelmachowicz, P.G., Beauchaine, A.K., and Jesteadt, W., "Normative thresholds in the 8- to 20-kHz range as a function of age", J.Acoust.Soc.Am., *86*, 1989, 1384-1391.
10. Makous, J.C. and Middlebrooks, J.C., "Two-dimensional sound localization by human listeners", J.Acoust.Soc.Am., *87*,1990, 2188-2200.
11. Oldfield, S.R. and Parker, S.P.A., "Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions", Perception, *13*, 1984, 581-600.
12. Middlebrooks, J.C. and Green, D.M., "Sound localization by human listeners", Ann.Rev. Psychol., *42*, 1991, 135-59.

Passive and Active Techniques for Hearing Protection

A. Dancer, K. Buck, P. Hamery and G. Parmentier
French-German Research Institute of Saint-Louis
5 rue du général Cassagnou, BP 34
68301 SAINT-LOUIS Cedex
France

SUMMARY

The present state of passive and active techniques for hearing protection in the military environment is reviewed. Solutions which allow to protect the ear while preserving the operational abilities of the personnel (detection, localization, communication...) are especially emphasized.

1. INTRODUCTION

In the military environment, personnel can be exposed to very high-level noises: impulse noises produced by weapons can reach 190 dB peak at the ear, continuous noises in the vicinity of jet engines can easily exceed 130 dBA! Although these extreme exposures conditions are relatively infrequent and concern only a few people, they present serious problems as they can produce immediate cochlear lesions and hence, large Permanent Threshold Shifts (PTS) [1]. Moreover, even "moderate" intensity noises: impulse noises of 150 to 165 dB peak (such as those produced by rifles in military training) [2], continuous noises of 100 to 120 dBA (such as those existing in armored vehicles or planes), are well over the admissible exposure conditions (i.e., ISO 1999) [3]. Altogether these noises correspond to a major cause of acoustic trauma among the military personnel [4,5]: in 1995, the US spent about 250 million dollar to compensate for the Noise Induced Hearing Losses (NIHL) of 59,088 veterans!

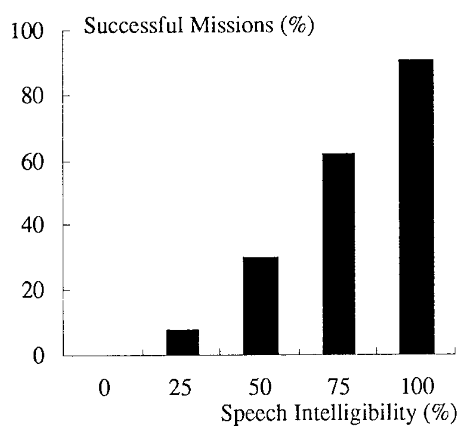


Fig. 1. Tank performance: percentage of successful missions (including navigation, reporting and gunnery), as a function of speech intelligibility (after Peters and Garinther [7])

Personnel exposed to weapon noises must be equipped with correctly fitted Hearing Protectors (HP) offering an adequate performance. However, the use of HP (together with the pre-existing hearing conditions: TTS and/or PTS) induces difficulties to detect, localize and identify acoustic sources in the environment, discomfort, and impedes the efficiency and the security of the soldier [6]. The HP also generally reduce the speech intelligibility. Actually, the masking of the communications by noise is a complicated phenomenon which depends on the characteristics of the speech signal and of the interfering noise, as well as on the type of HP and of the intercom system. Decrease in speech intelligibility can drastically reduce the global performance of complex and expensive weapon systems [7] (fig. 1).

The HP must be designed and evaluated by taking into account as well the need for hearing protection as the consequences of their use for the operational abilities. Moreover, the HP should be fitted to each soldier station! If not, the risk is that the protection will not be worn. Therefore, a general-purpose device is not feasible. Keeping these requirements in mind, we shall review some aspects of the passive and active HP techniques.

2. NOISE REDUCTION AND INSERTION LOSS MEASUREMENT TECHNIQUES

Noise Reduction (NR) is the difference between the incident and the received sound pressure levels (SPLs) when a HP is worn (i.e., between the free field and the entrance to the earcanal - or the tympanum - for an earmuff; between the free field and the tympanum for an earplug). Insertion Loss (IL) is the difference between the SPLs measured at a reference point (i.e., at the entrance to the earcanal - or at the tympanum - for an earmuff; at the tympanum for an earplug) before and after a HP is put into place [8].

For the assessment of the attenuation afforded by earplugs and earmuffs at very high levels, the classical measurements performed by means of the subjective method: Real-Ear-At-Threshold (REAT) at low steady-state noise levels (according to ISO 4869-1 for example) [9] are not suitable. First of all, this method does not allow to evaluate the peak pressure of an impulse under a HP (the ISO standard 1999 [3] restricts the equal energy principle to peak levels below 140 dB, the American Conference of Governmental Industrial Hygienists [10] does not allow the unprotected exposure to impulses above 140 dB peak, and most Damage Risk Criteria - DRC - for weapon noises [1] take into account, besides the duration and the number of the impulses, the peak pressure [11,12]). Even if serious doubts exist about the pertinency of "peak level" measurements under HP as part of the classical DRC [13], it is nevertheless essential to get this information. Moreover, the behavior of the HP undergoing the action of high-level noises may exhibit nonlinearities. The apparition of a nonlinearity, its importance, the

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

variation of its characteristics as a function of the parameters of the noise as well as its net effect (either "positive" or "negative" as far as the global attenuation is concerned), are generally unpredictable. For this reason the attenuation of each HP should be measured in the actual exposure conditions for which it is intended to be used! The same kind of limitations apply to the "Microphone-In-Real-Ear" (MIRE) measurement technique [14]. Moreover, this technique is presently unsuitable for earplugs attenuation measurements and, last but not least, MIRE is impossible to use as a routine technique with high-intensity impulses because of the security of the subjects.
Therefore, the only possibility to assess the actual behaviour of the HP when exposed to high-level (impulse or continuous) noises, to characterize their nonlinearity (if any), and to measure amplitude spectrum and peak pressure attenuations, is to use an Artificial-Test-Fixture (ATF) and preferably an artificial head with an ear simulator [15,16].

ATF are currently used to evaluate the physical attenuation afforded by earmuffs in steady-state noise. In these conditions the ATF must comply with standards (ISO and/or ANSI standards) [17,18]. However, ATF which are commercially available (Brüel & Kjaer, Knowles Electronics Manikin for Acoustic Research, Head Acoustics GmbH) are either not suited for our purpose and/or present a poor acoustic isolation (fig. 2).

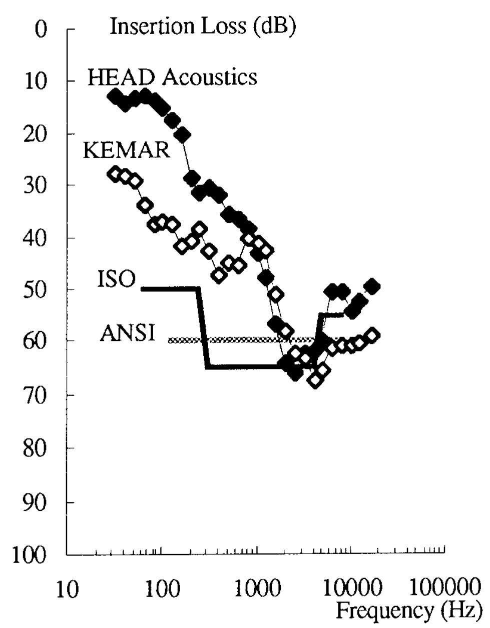


Fig. 2. Insertion Loss performances of the KEMAR and HEAD Acoustics (first version) ATFs
Minimal Insertion Loss of ATF (in dB) as defined by the ISO and ANSI standards (1/3 oct. bands)

We designed a new ATF in order to reach better performances. The "head" was arranged to fit: (i) the HEAD Acoustics GmbH device corresponding to the external ear and the circumaural region, (ii) the Brüel & Kjaer ear simulator (type 4157). To allow the measurement of peak levels up to 190 dB, the 1/2" Brüel & Kjaer microphone (type 4134) of the original ear simulator is replaced by an underpolarized (28 V instead of 200 V) 1/4" microphone (type 4136) (fig. 3).

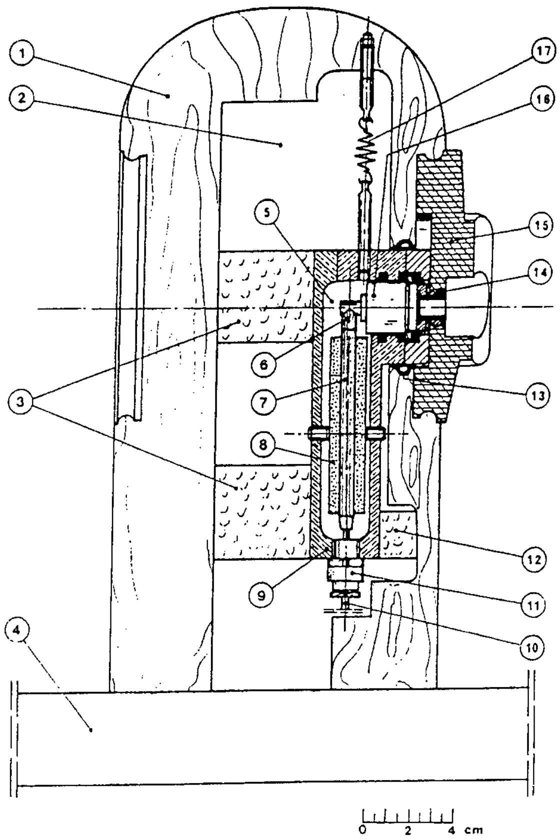


Fig. 3. Cross section of the ATF
1. shell
2. head cavity
3. damping foam blocks
4. wood base
5. cavity of the brass shell surrounding the measuring equipment
6. B&K bend (type WU 0278)
7. microphon preamplifier (B&K 2633)
8. foam sleeve
9. brass shell
10. cable
11. stuffing box
12. damping foam block
13. circular coupling
14. auditory canal extension (HEAD Acoustics)
15. outer ear cheek (HEAD Acoustics)
16. ear simulator (B&K 4157) equipped with a B&K microphon (type 4136) and a B&K adaptator (1/2" - 1/4")
17. suspending spring

The measured Transfer Function of the Open Ear (TFOE) of this ATF is in close agreement with the experimental data published by Shaw [19] and can be considered as linear up to a peak pressure of about 190

dB at the ear simulator microphon. Thanks to a special assembling design and to various suspending and damping devices, when the ATF "earcanal" is closed the maximum IL afforded by the ATF is better than 80 dB from 0.4 to 5 kHz (fig. 4) and well over the ANSI and ISO criteria.

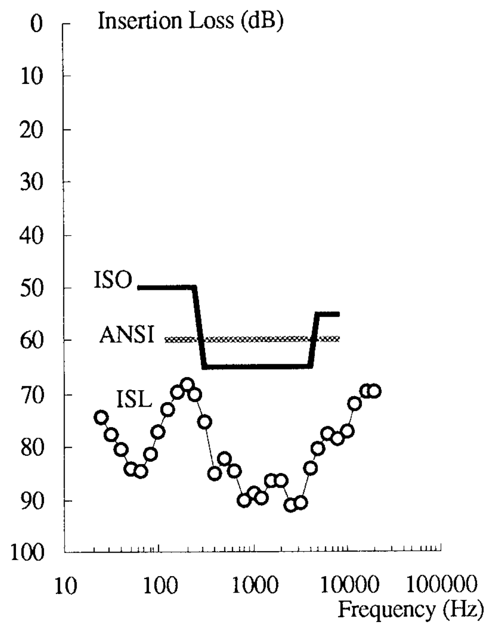


Fig. 4. Insertion Loss performances of the ATF (1/3 oct. bands) compared to ISO and ANSI standards

It is then possible to study the IL of Hearing Protectors without any limitation (≥ 80 dB) of the dynamics of the measurements! Generally speaking such large IL values are not taken into account for hearing protection evaluation because they exceed the Bone Conduction (BC) thresholds [20,21]. However, the measurements which are feasible thanks to the large dynamics of our ATF allow (i) to determine the physical performance of (almost) any HP, (ii) to know the actual pressure-time history existing below a HP under (almost) any exposure conditions, and (iii) to apply any correction curve corresponding to either the BC limits, or to the Physiological Masking noise (PM) and the Occlusion Effect (OE) (see for example: Schroeter and Poesselt [20]), and thus allow a very general approach of the measurement of the HP efficiency. Moreover, it must be noted that all investigations concerning the BC limits rely on threshold detection methods and that no proof whatsoever exists to indicate that the BC NIHL are comparable to those produced by aerial conduction!

This ATF is perfectly suitable for measurements with earmuffs. Concerning the earplugs, although the HEAD Acoustics device provides some simulation of the "earcanal" tissues this point still needs to be improved: i.e., thickness, compliance, geometry... and requires international standardization. Moreover, as the mechanical behavior of some earplugs (i.e., foam plugs) depends largely on their temperature, it will be necessary to control the temperature of the "earcanal" wall and to stabilize it around its physiological value.

3. ASSESSMENT OF THE HEARING PROTECTION EFFICIENCY

There are two main methods to decide whether the hearing protection afforded by a HP is sufficient: either (i) by measuring the signal close to the head of the subject and using the IL characteristics of the protector in order to calculate the equivalent dose of acoustic energy to which the subject would be exposed unprotected [8], or (ii) by measuring the pressure-time signature under the protector and introducing the peak pressure and duration into the classical criteria for weapon noises. Actually the second possibility, which is sometimes used as far as impulse noises are concerned, consists of an untested extension of the use of the weapon noise criteria (DRC) because they have been primarily designed to apply to pressure-time signatures measured on the outside and to unprotected ears. In some other instances, a global protection factor has been applied to the peak pressure before its evaluation by the criterion. This method is incorrect too because there is no direct relation between the peak pressure attenuation afforded by a protector and its Insertion Loss characteristics.

Which method is the most representative of the actual hearing protection afforded by a protector? The peak pressure attenuation grossly underestimates the protection afforded by the HP when used in conjunction with the classical DRC. Actually, the risk corresponding to the exposure to a slow rise time signal (as recorded under a HP) is in fact much lower than the risk corresponding to the exposure to a shock wave with an instantaneous rise time (with the same peak pressure). The LAeq8 attenuation based on IL measurements gives in most exposure conditions a good evaluation of the auditory hazard but in some other cases it still underestimates the efficiency of the HP. A still better prediction could be achieved by using a weighting function corresponding to the curve of hearing sensitivity at threshold instead of the A-weighting function. On the other hand, it could well be that the very high protection afforded by ordinary hearing protectors (standard earmuffs are able to fully protect the ear against 100 impulses of 187 dB peak pressure simulating artillery noise [22,23,24]) is due to the nonlinearity of the middle ear. Price and Kalb [25,26,27] emphasized the limitation of the tympano-ossicular chain displacements due to the nonlinear mechanical characteristics of the annular ligament when exposed to large impulses. If important for unprotected exposures, this effect could be essential in order to understand the surprisingly small damages induced by large but slow-rising impulses as those existing under HP.

4. INSERTION LOSS OF HEARING PROTECTORS IN HIGH LEVEL IMPULSE NOISE

The attenuation afforded by earplugs (E.A.R. foam, E.A.R. Ultrafit, E.A.R. Ultratech, E.A.R. Link, RACAL Airsoft, RACAL Gunfender, perforated earplugs...), and earmuffs (WILLSON SB 258, E.A.R Ultra 9000...) was measured during exposure to Friedlander waves of ≈ 150, 170 and 190 dB peak

pressures (A-durations: ≈ 0.2 and ≈ 2 ms) under normal and grazing incidences [16]. Typical Pressure-Time history as well as amplitude spectra of these impulses are presented on figure 5.

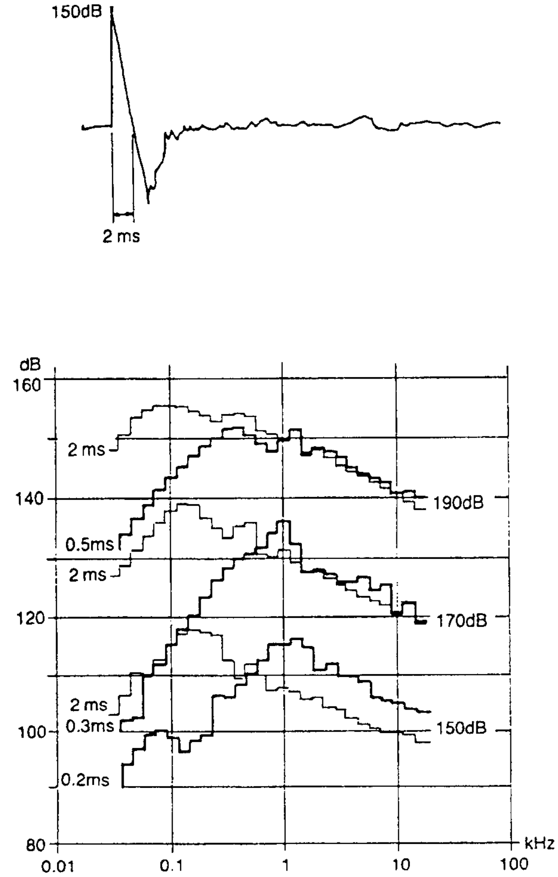


Fig. 5. Pressure time signature of a Friedlander wave (peak pressure: 150 dB, A-duration: 2 ms) and 1/3 oct. band amplitude spectra of typical impulses (peak pressure: 150, 170 and 190 dB peak pressure, A-durations: 0.2 to 2 ms)

Some HP behave linearly: no significant modification of the IL is observed when the peak pressure of the impulse changes. Figure 6 presents the IL of the RACAL Airsoft earplug as a function of frequency (1/3 octave bands) for impulses of ≈ 150, 170 and 190 dB peak (≈ 2 ms A-duration, normal incidence). It is interesting to notice that the attenuation of the RACAL Airsoft earplug as measured by REAT methods (ISO 4869 and ANSI S 3.19-1974) at low steady-state noise levels by the Berufsgenossenschaftliches Institut für Arbeitssicherheit [28] is comparable to our results obtained with high-level impulses. This indicates clearly (i) that this earplug behaves linearly and (ii) that the ATF reproduces reasonably well the average behavior of the human ear.

Some HP behave nonlinearly. In some instances the nonlinearity is unfavourable: the higher the peak pressure, the lower the IL (fig. 7), in some other instances the nonlinearity is favourable: the higher the peak pressure of the impulses, the higher the IL (fig. 8 and 9). From these examples, it is particularly obvious that the actual attenuation provided at high levels cannot be inferred from REAT attenuation values.

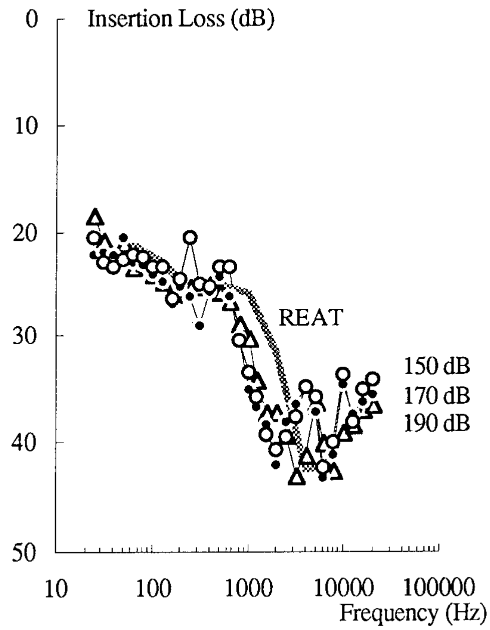


Fig. 6 Insertion Loss afforded by the RACAL Airsoft earplug for different peak pressure levels of the impulses 150 dB, 170 dB and 190 dB (A-duration: 2 ms, normal incidence)
REAT Insertion Loss measured by B.I.A. [28]
(1/3 oct. bands)

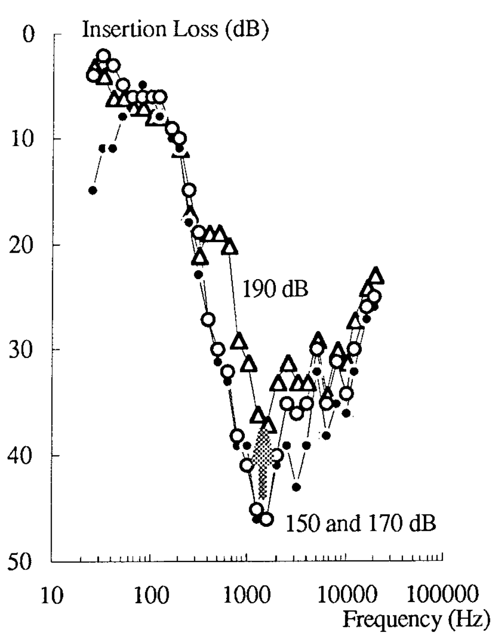


Fig. 7. Insertion Loss afforded by the WILLSON SB 258 earmuff for different peak pressure levels of the impulses: 150 dB, 170 dB and 190 dB (A-duration: 2 ms, normal incidence) (1/3 oct. bands)

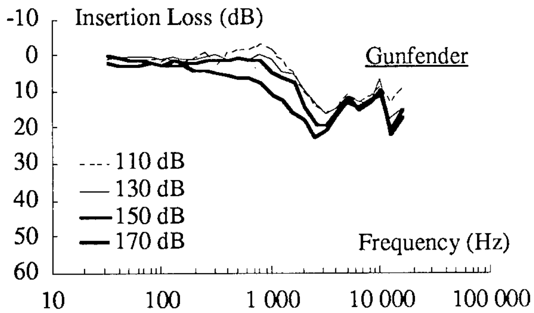


Fig. 8. Insertion Loss afforded by the RACAL Gunfender earplug for different peak pressure levels of the impulses: 110 dB, 130 dB, 150 dB and 170 dB (A-duration: 2 ms, normal incidence)

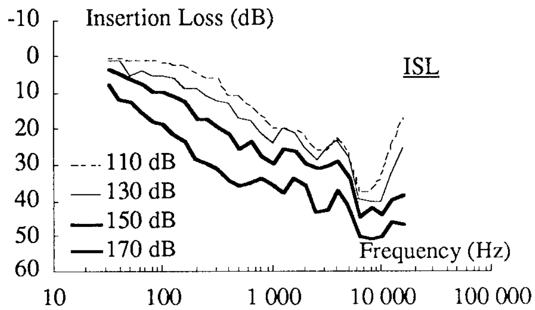


Fig. 9. Insertion Loss afforded by the ISL earplug for different peak pressure levels of the impulses: 110 dB, 130 dB, 150 dB and 170 dB (A-duration: 2 ms, normal incidence)

4. NONLINEAR EARPLUGS

The perforated earplugs present an attenuation which increases with the peak pressure (acoustic resistance through the orifice(s) increases with the peak level). The former nonlinear plugs (Gunfender, fig. 8) acted nonlinearly only beyond 140 dB and the IL increased by about 5-10 dB for each 20 dB increase of the peak pressure of the impulse (NR peak attenuation increased by 10-15 dB from 130 to 190 dB: fig. 10) [29,30]. New designs by ISL [31,32] allow the nonlinearity to begin around 110 dB (i.e., below the potentially dangerous levels for this kind of impulses), and to get very large IL values for the highest peak pressures (fig. 9) (NR peak attenuation increases by 20-25 dB from 110 to 190 dB: fig. 10).

Actually, to protect the ear against impulse noises, nonlinear perforated earplugs are a very attractive solution. They are light, cheap, easy to clean and to maintain, they work without any energy supply and without intervention of the subject, and are compatible with other head equipments. Unlike classical plugs and because they present a limited insertion loss at the low and moderate levels (especially at low frequencies where it can even be zero up to 1 kHz), these plugs allow speech communication, detection and localization of acoustic sources in about the same conditions as an unprotected subject (fig. 11), thus avoiding overprotection problems! All the same, they afford a protection adapted to occasional exposures to impulse noises such as those produced during training or combat [33]. Moreover, recent human studies by Johnson et al. (unpublished data) have demonstrated that these new earplugs are efficient (no significant TTS) for repeated exposures up to 187 dB peak (Friedlander waves, 1.5 ms A-duration, free field, normal incidence, 100 rounds), when they are well fitted.

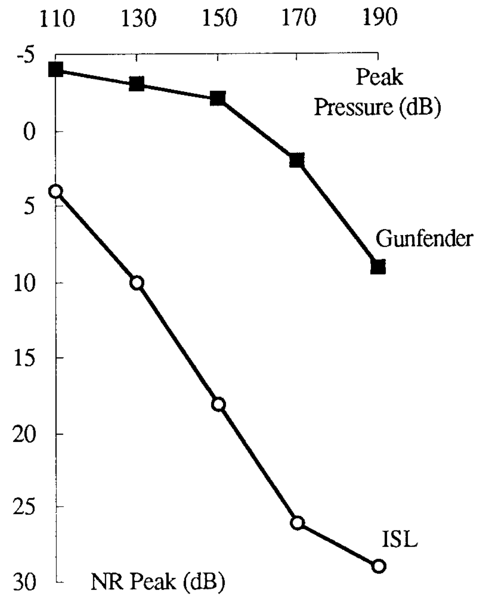


Fig. 10. Peak attenuation (NR values) afforded by the nonlinear RACAL Gunfender earplug and the ISL earplug as a function of the peak pressure of the impulses (Friedlander waves)

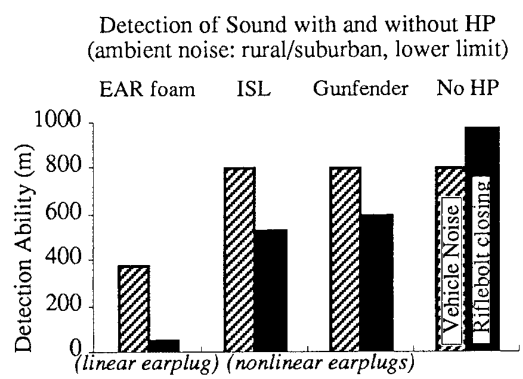


Fig. 11. Detection ability (in meters) of various sounds (vehicle noise, riflebolt closing) with and without hearing protectors (linear and non linear earplugs) in low level ambient noise (after Garinther et al., [6])

Some work is presently in progress in order to determine the best design to ensure a good and easy seal, as well as a good comfort when the earplug is worn during a long time. A good seal and a very satisfactory

comfort index can be achieved with the help of custom molded earplugs. The higher cost and the making delay of these plugs is obviously a major drawback but their use by long term volunteers might be favourably considered. Finally, to allow a protection against continuous noise (during transportation in APC for example), a miniature rotating valve can be added to the earplug design to improve the passive attenation charcteristics at "moderate" levels.

5. DOUBLE HEARING PROTECTION AND SPEECH INTELLIGIBILITY

The efficiency of a double protection is usually limited at the low frequencies by the coupling between the protectors and by the compliance of the skin of the external auditory canal and, at the high frequencies, by bone conduction [20]. According to Gierke [34], a double protection improves only by 10 dB on an average the IL from 125 to 8000 Hz. As a good protection can be afforded by (well-fitted) ordinary HP, the use of a double hearing protection (earmuff and earplugs) is not necessary in case of exposure to impulse noise [24,33]. Actually, the main interest of the double hearing protection is to protect the ear against loud continuous noise and to improve speech intelligibility.

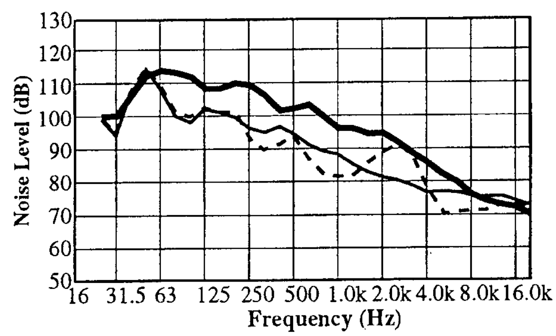


Fig. 12. Noise level (1/3 oct. bands) into an armored vehicle: 121 dB SPL, 108 dBA (solid thick line); into a turbopropeller plane : 117 dB SPL, 99.2 dBA (solid thin line); under the pilot's headset ot the turbopropeller plane with the voice communication system "ON": 117 dB SPL, 99.7 dBA (dashed line)

Levels as high as 120-125 dBA can be measured in armoured vehicles (100-110 dBA in turbopropeller planes) (fig. 12). As these noises contain a lot a low frequencies, a simple earmuff is not enough to achieve the protection of the ear [28]: inside the Bradley vehicle the noise is about 114 dBA and the sound level at ear is 100 dBA with the current tanker helmet, therefore allowing a daily exposure time of 20 minutes only [38]! Actually, most of the current earmuffs/headsets present a very poor attenuation at low frequencies. On figure 13 we can compare the passive attenuation afforded by two military headsets used in armoured vehicles: LE 132 (former French HP) and BOSE CVC (new US HP) when measured by the ATF under the same experimental conditions. The newly designed HP (BOSE CVC) is much better: on the low frequency side, this improvement is mainly due to a new type of seal. However, we must note that these attenuation curves correspond to maximal values. When used on man, the attenuation values will drop because of fitting problems and of the possible use of glasses, MOPP. Therefore, in practice the low frequencies will go through the muff with very little attenuation and mask the speech signals.

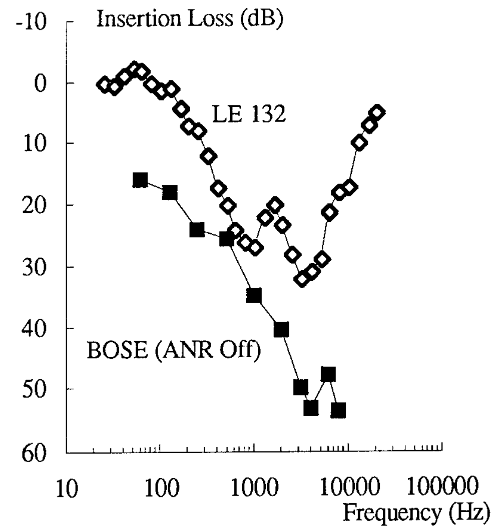


Fig. 13. Passive attenuation afforded by two tanker helmets

The masking ruins the performances of the communication system [35] and forces the subject to increase the speech level. It is then possible to measure a larger level under the muff than in the vehicle itself when the communication system is keyed (fig. 12)! Under these conditions the speech intelligibility is no more determined by the global attenuation of the HP but by its performances at the low frequencies only. The low frequencies mask medium and high frequencies and this masking effect is nonlinear (the larger the level, the larger the spread and the amplitude of the masking) (fig. 14).

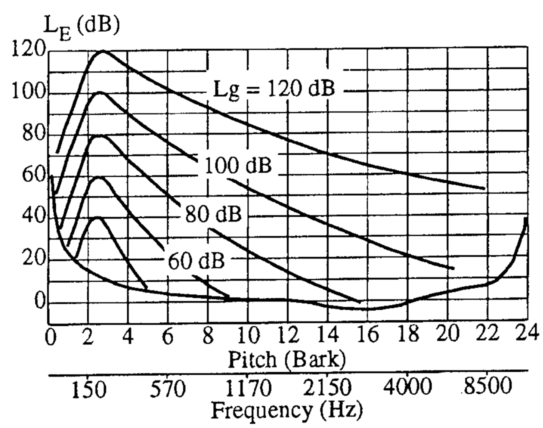


Fig. 14. Masking curves corresponding to a "critical band" noise at 250 Hz (LE: excitation level, Lg: global level) (after Zwicker and Feldkeller, 1967)

Figure 15 allows to better understand this phenomenon. We can observe that it is not the attenuation performance of the HP at medium frequencies which determines the speech level as adjusted by the listener, but the masking effect due to the very low frequencies. A better HP, as far as the medium and high frequencies are concerned, would not perform better with respect to the communications. The only solution is to improve the low frequency attenuation characteristics of the HP. As discussed before, it is possible to increase this attenuation but the actual performances of the system in the real life will always come short from the values measured in the laboratory.

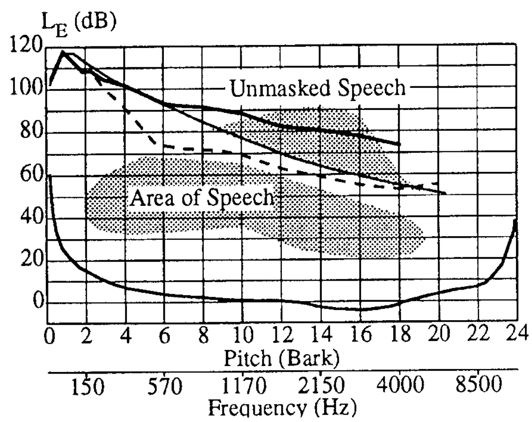


Fig. 15. Noise level (1/3 oct. bands) into a turbopropeller plane (solid thick line); noise level under the pilot's headset (dashed line), masking curve (solid thin line)

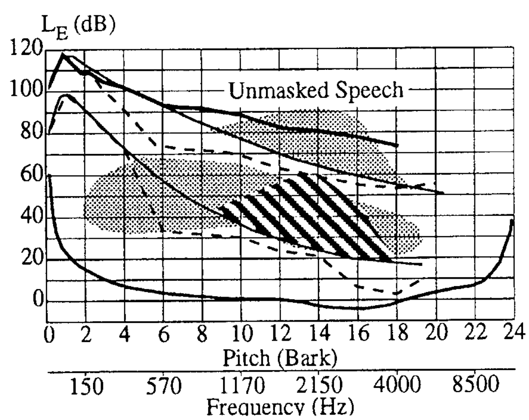


Fig. 16. Noise level (1/3 oct. bands) into a turbopropeller plane (solid thick line); noise level at the pilot's ear: (i) with a single hearing protection, (ii) with a double hearing protection (dashed lines), masking curves (solid thin lines)

In these circumstances, the use of a double hearing protection was recommended by Kryter a long time ago [36,37]. The level difference between the speech signal and the noise is not modified over the whole frequency range but the earplugs worn under the earmuff bring both signals (noise and speech) down to a level at which the ear is not saturated and at which the masking by the low frequencies is less effective. On figure 16 we can see that it is always the masking effect due to the very low frequencies which determines the speech level but (i) this effect is smaller and (ii) the overall level at ear is reduced. We have shown (Dancer and Pellieux, unpublished data) that in this configuration the same speech intelligibility (CVC test) can be achieved with a reduction of 20 dB of the global level (speech + noise) at the level of the ear, therefore allowing a long term exposure and an improvement of the performances. This very simple and cheap method is well received by the subjects and gives results with are comparable to those obtained with the help of the Active Noise Reduction (ANR) techniques.

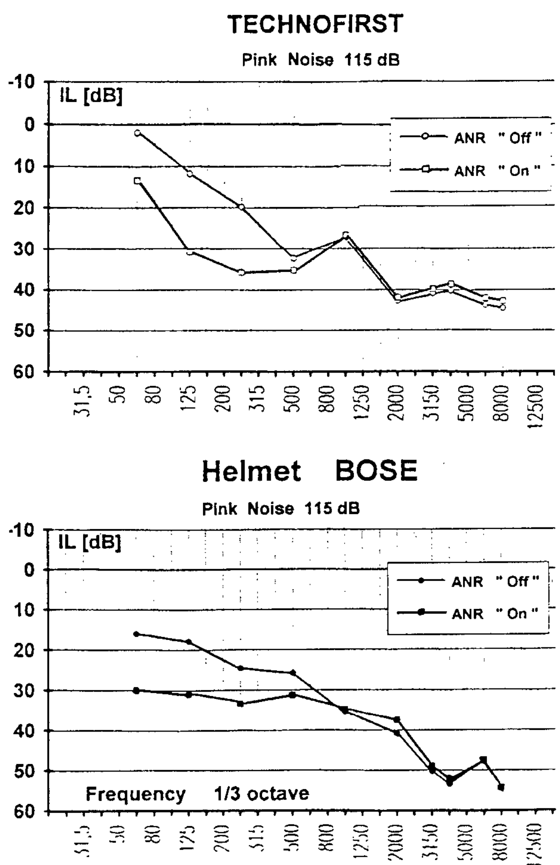


Fig. 17. "Passive" and "Active + Passive" attenuations afforded by two ANR earmuffs

6. ANR TECHNIQUES AND SPEECH INTELLIGIBILITY

For the same reasons (limited attenuation at the low frequencies by the ordinary HP and speech masking problems), ANR devices are used in the military environment [38,39]. Nixon et al. [40] summarized the main characteristics and possible applications of ANR. The present ANR earmuffs improve the attenuation by a maximum of 20 dB at the low frequencies (between 50 and 300 Hz) and then get an insertion loss comparable to a double hearing protection in this frequency range (fig. 17). It is interesting to note that the HP which offer the best overall attenuation are not those with the

active noise reduction! This emphasizes once more the importance of the passive characteristics of the HP at the low frequencies. Figure 18 shows that in these circumstances and unlike what was observed with the double hearing protection, the characteristics of the HP at medium and high frequencies is the limiting factor for the speech intelligibility (the masking curve is then at a lower level). As a conclusion one should say that a good ANR HP must be first a very passive HP!

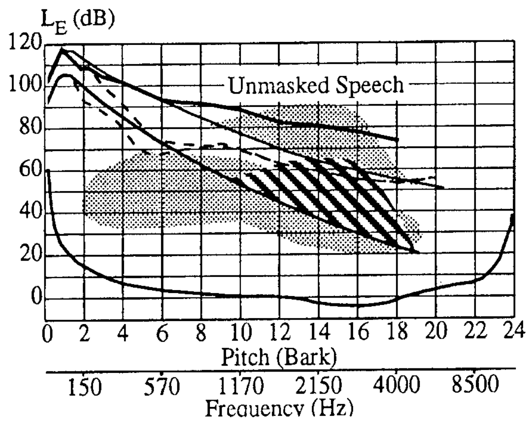


Fig. 18. Noise level (1/3 oct. bands) into a turbopropeller plane (solid thick line); noise level at the pilot's ear: (i) with a single passive hearing protection, (ii) with a single active hearing protection (dashed lines), masking curves (solid thin lines)

ANR earplugs which will be available in the future operate up to 2.5 kHz and represent a significative improvement of the system especially for the intelligibility of speech (which is only marginally improved by the present ANR earmuffs at least because of their only ANR performances).

Most of the ANR-HP work as well for continuous as for impulse noise. However, their efficiency is limited by the output level of the electro-acoustic system (120 - 130 dB) and they are of little use for large impulses [41].

Last but not least, the use of digital filtering could help to adjust the ANR systems to each user and/or to each noise exposure condition. ANR-HP would then be compatible with high quality speech intercom system (adaptative critical band filtering, neural networking) and even include a three-dimensional virtual reality [42]. Up to now, only tank crews, helicopter and fighter pilots benefit of the ANR, but this technique which will be soon an integrated part of the much more sophisticated and comprehensive head equipment of the soldier.

7. CONCLUSION

We hope that this paper will draw the attention on the very specific problems of noise and hearing protection in military life, avoid misinterpretations and/or negligences which are the cause of many hearing damages, and indicate some solutions to reduce greatly the noise hazard while preserving the operational capabilities of the soldier.

REFERENCES

1. Research Study Group on the Effects of Impulse Noise, "Effects of Impulse Noise", NATO Document 1987; AC/243, Panel 8/RSG.6, D/9, pp 1-84.

2. Pääkkönen, R., Kyttälä, I., "Attenuation of rifle and pistol noise with the use of suppressors and subsonic bullets", Acta Acustica, 2, 1994, pp 29-36.

3. ISO 1999, "Acoustics - Determination of Occupational noise exposure and estimation of noise induced hearing impairment", International Organization for Standardization, Geneva, 1990.

4. Borchgrevink, H.M., "One third of 18 year old male conscripts show noise induced hearing loss", in "Noise as a Public Health Problem", Swedish council for building research, 2, 1988, pp 27-32.

5. Ohlin, D., "Hearing loss disability trends and procedures", in DGA/DSTI, Etablissement Technique de Bourges, France (ed): Actes des Journées des 7 et 8 juin 1995 du Comité "Bruits d'Armes". 1995.

6. Garinther, G., Kalb, J., Hodge, D., Price, G., "Proposed aural nondetectability limits for army materiel", US Army Human Eng. Lab 1985, Techn. Memorandum, pp 3-85.

7. Peters, L., Garinther, G., "The effects of speech intelligibility on crew performance in an M1A1 tank simulator", US Army Human Eng. Lab, 1990, Techn. Memorandum, pp 11-90.

8. Berger, E.H., "Methods of measuring the attenuation of hearing protection devices", J. Acoust. Soc. Am., 79, 1986; pp 1655-1687.

9. ISO 4869-1, "Subjective method for the measurement of sound attenuation", International Organization for Standardization, Geneva, 1990.

10. American Conference of Governmental Industrial Hygienists, "TLV's threshold limit values and biological exposures indices for 1988-1989", Cincinnati, 1988.

11. Pekkarinen, J.O., Starck, J.P., Ylikoski, J.S., "Hearing protection against high-level shooting impulses in relation to hearing damage risk criteria", J. Acoust. Soc. Am., 91, 1992, pp196-202.

12. Dancer, A.L., "Status and Shortcomings of Military Noise Standards", in "Noise-Induced Hearing Loss", Dancer, A.L., Henderson, D., Salvi, R.J., Hamernik, R.P., eds., Saint-Louis, Mosby Year Book, 1992, pp 513-520.

13. Dancer, A.L., Franke, R., "Hearing Hazard from Impulse Noise: a comparative study of Two Classical Criteria for Weapon Noises (Pfander Criterion and Smoorenburg Criterion) and the LAeq8 Method", Acta Acustica, 3, 1995, pp 539-547.

14. Hellström, P., "Objective methods for evaluating conventional, nonlinear, and active hearing protector attenuation, in "Noise-Induced Hearing Loss", Dancer, A.L., Henderson, D., Salvi, R.J., Hamernik, R.P., eds., Saint-Louis, Mosby Year Book, 1992, pp 401-412.

15. Franke, R., Parmentier, G., Buck, K., Evrard, G., Kronenberger, G., Beck, C., Brunner, J.J., "Tête artificielle pour l'évaluation de l'efficacité des protecteurs auditifs vis-à-vis de bruits de niveau élevé, I. Développement et essais en régime impulsionnel", ISL Report 112/1994.

16. Parmentier, G., Franke, R., Buck, K., Kronenberger, G., Evrard, G., "Tête artificielle pour l'évaluation de l'efficacité des protecteurs auditifs vis-à-vis de bruits de niveau élevé. II. Application à différents types de bouchons d'oreilles et de serre-tête", ISL Report 113/1994.

17. ISO TR 4869-3: 1989 (E), "Simplified method for the measurement of insertion loss of ear-muff type protectors for quality inspection purposes", International Organization for Standardization, Geneva, 1989.

18. ANSI S3.19-1974, "Method for the measurement of the real ear protection and physical attenuation of earmuffs", American National Standards Institute, New York, 1974.

19. Shaw, E.A.G., "The external ear", in "Handbook of Sensory Physiology/Auditory System" Keidel, W.D., Neff, W.D., eds., Berlin, Springer Verlag, vol. 1, 1974, pp 455-490.

20. Schroeter, J., Poesselt, C.H., "The use of acoustical test fixtures for the measurement of hearing protector attenuation. Part II: Modeling the external ear, simulating bone conduction, and comparing test fixture and real ear data", J. Acoust. Soc. Am., 80, 1986, pp 505-527.

21. Berger, E.H., "Laboratory attenuation of earmuffs and earplugs both singly and in combination", Am. Ind. Hyg. Assoc. J., 44(5), 1983, pp 321-329.

22. Johnson, D.L., Patterson, J.H., "Rating of hearing protector performance for impulse noise", in "Proceedings Hearing Conservation Conference", University of Kentucky ed., 1992, pp 103-106.

23. Patterson, J.H., Mozo, B.T., Johnson, D.L., "Actual effectiveness of hearing protection in high level impulse noise", in "Noise as a public health problem", Vallet, M., ed., Bron, France, INRETS, vol. 3, 1993, pp 122-127.

24. Patterson, J.H., Johnson, D.L., "Temporary Threshold Shifts produced bu High-Intensity Free-Field Impulse Noise in Humans Wearing Hearing Protection", in "Scientific Basis of Noise-Induced Hearing Loss", Axelsson, A., Borchgrevink, H., Hamernik, R.P., Hellstrom, P-A., Henderson, D., Salvi, R.J., eds.,. Thieme, New York, 1996, pp 313-320.

25. Price, G.R., "An upper limit to stapes displacement: implications for hearing loss", J. Acoust. Soc. Am., 56, 1974, pp 195-197.

26. Price, G.R., Kalb, J.T., "Insights into hazard from intense impulses from a mathematical model of the ear", J. Acoust. Soc. Am., 90, 1991, pp 219-227.

27. Price, G.R., "Modeling the ear's response to intense impulses and the development of improved damage risk criteria", in "Proceedings Hearing Conservation Conference", University of Kentucky ed., 1992, pp 93-96.

28. Kusy, A., "Efficacité et confort des protecteurs individuels contre le bruit", Travail et Sécurité (INRS), 1991, pp 215-243.

29. Forrest, M.R., Coles, R.R.A., "Problems of communication and ear protection in the royal marines", J. R. Nav. Med. Serv., 56, 1970, pp 162-169.

30. Shaw, E.A.G., "Hearing protector design concepts and performance limitations", in "Personal Hearing Protection in Industry", Alberti, P.W., ed., New York, Raven, 1982, pp 51-68.

31. Dancer, A., Cabanis, A., Grateau, P., Lafont, D., Pfannstiel, E., "Interest of earplugs: field studies and new prototype", French-German Research Institute of Saint-Louis, France, ISL Rpt PU 345/95, 1995.

32. Albe, P., Evrard, G., Dancer, A., "Mise au point de bouchons d'oreilles passifs non linéaires", French-German Research Institute of Saint-Louis, France, ISL Rpt R 102/96, 1996.

33. Dancer, A., Grateau, P., Cabanis, A., Barnabé, G., Cagnin, G., Vaillant, T., Lafont, D., "Effectiveness of earplugs in high-intensity impulse noise", J. Acoust. Soc. Am., 91, 1992, pp 1677-1689.

34. Gierke, H.E., "Personal protection", Noise Control, 2(1), 1956; pp 37.

35. Price, G.R., "Weapon noise and performance", in Etablissement Technique de Bourges, France, CR n°241/CT/DTB/1995, 1995

36. Kryter, K.D., "Effects of ear protective devices on the intelligibility of speech in noise", J. Acoust. Soc. Am., 18, 1946, pp 413-417.

37. Kryter, K.D., "Speech Communication in Noise", Operational Applications Laboratory, Air Force Cambridge Research Center, Bolling Air Force Base, Washington, D.C., Rpt. AFCRC TR 54-52, 1955.

38. Anderson, B.W., Garinther, G., "ANR interest and applications: I. The state of art in the US Army", in Etablissement Technique de Bourges, Bourges, France, CR n°241/CT/DTB/1995, 1995.

39. McKinley, R.L., Steuver, J.W., Nixon, Ch.W., "Estimated reductions in Noise-Induced Hearing Loss by application of ANR headsets", in "Scientific Basis of Noise-Induced Hearing Loss", Axelsson, A., Borchgrevink, H., Hamernik, R.P., Hellstrom, P-A., Henderson, D., Salvi, R.J., eds., Thieme, New York, 1996, pp 347-360.

40. Nixon, Ch.W., McKinley, R.L., Steuver, J.W., McCavitt, A.R., "What can active noise reduction headsets do for you", in "Proceedings of the Hearing Conservation Conference", university of Kentucky ed., Cincinnati, University of Kentucky, BU156, 1992, pp 107-110.

41. Dancer, A., Franke, R., Parmentier, G., Buck, K., "Actual performance of hearing protectors in impulse noise/nonlinear behaviour", French-German Research Institute of Saint-Louis, France, Rpt. ISL PU-319/94, 1994.

42. Pellieux, L., Piedecocq, B., Quandieu, P., "Stereophonic presentation of information: space simulation under earphones", Journal de Physique IV, Colloque C1, Suppl III, 2, 1992, pp 395-398.

Active Hearing Protectors new Developments and Measurement Procedures

K. Buck[*] and G. Parmentier[**]
* Group MMI
** Group APS
French-German Research Institute
3, rue du Général Cassagnou
BP 34
F 68301 Saint-Louis CEDEX
France

1. SUMMARY

The need for active noise cancellation (ANC) hearing protectors in the armed forces is shown. A description of the systems that are actually commercially available and of the way that future systems may be designed is described. It is also presented, that the presently normalized evaluation procedures should be modified to suit better the new technology of active hearing protectors.

2. INTRODUCTION

A typical noise, that is encountered in an armored vehicle (figure 1) or in propeller driven airplanes, is characterized by very high levels (up to 120 dB) that are concentrated at very low frequencies (between 50 and

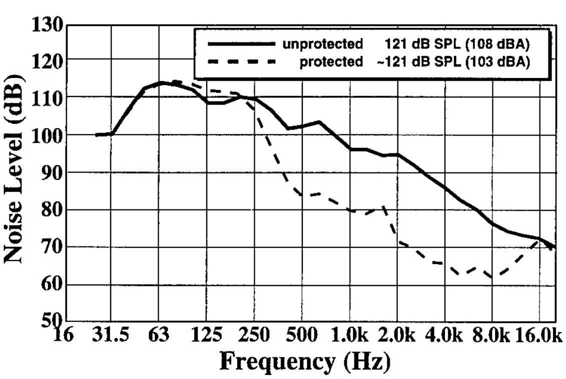


Fig. 1: Third octave band Levels inside an armored vehicle (AMX 30), with and without the standard hearing protection for Frenc tank crews.

250 Hz). Crew members are not allowed to an exposure of more than some minutes if unprotected. The standard hearing protectors used in the army are not effective for low frequencies, and so even a soldier wearing hearing protectors will not be allowed to a daily exposure of more than about 20 minutes. This limitation in exposure time doesn't allow meaningful training on the weapon system, and so may lead to reduced effectiveness.

Another factor that may lead to a reduced effectiveness of the whole weapon system, is the reduction of speech intelligibility due to masking of the speech area by the low frequencies of the spectrum. These two facts, that may introduce a decrease of performance of the soldier and the whole weapon system, can only be solved if the level of the low frequency noise is reduced. One possibility for reducing this components of the noise, is the use of active hearing protectors.

3. PRINCIPLE OF ACTIVE HEARING PROTECTION

The principle of active hearing protection has been patented by LUEG in 1936 [1]. The principle has also been described by JESSEL and ANGEVINE in 1980 [2], by LEVENTHALL in 1981 [3] and FRANKE et col. in 1986 [4]. The principle of active noise control is

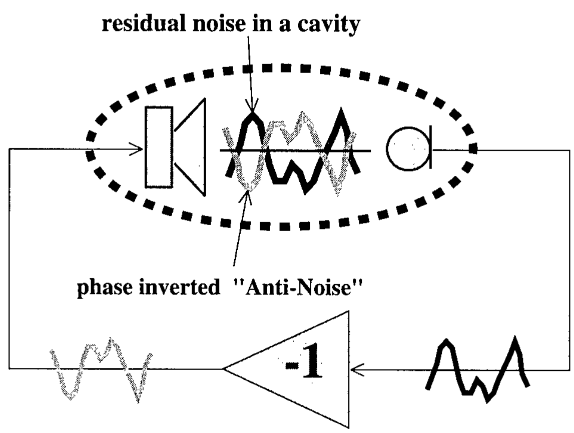
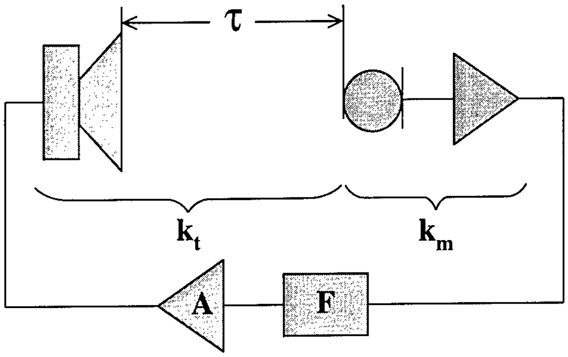


Fig. 2: Basic principle of an ANC Device

shown in figure 2. It consists in measuring the residual noise inside a cavity (e.g. the noise under earmuff), this noise is phase inverted and fed back with a loudspeaker into the same cavity. The two acoustic pressures (the measured one and the inverted) then will theoretically add to 0. A system as simple as this however, would not be stable in reality. Therefore the feedback loop of an actual ANC (Active Noise Cancellation) system (fig. 3) includes a compensation filter. When designing this

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

filter, we have to take into account the different transfer functions that are shown in figure 3. The



A linear amplifier
F feedback compensation filter
k_t transfer function of loudspeaker + volume
k_t transfer function of microphone + preamplifier
k_t, k_m and F are frequency dependent

Fig. 3: Feedback loop of an ANC system

formula 2 below shows, that the closed loop system becomes unstable if the open loop gain (1) becomes -1. The transfer function of the compensation filter has

open loop gain $$F\,A\,K_t\,K_m \quad (1)$$

closed loop Gain $$\frac{1}{1 + F\,A\,K_t\,K_m} \quad (2)$$

active attenuation $$20 \log_{10} |1 + F\,A\,K_t\,K_m| \quad (3)$$

now to be determined in a way, that the system is stable. If the system is stable, the active attenuation is expressed by (3).

4. COMMERCIALIZED ANC HEARING PROTECTORS

Presently about 10 different ANC hearing protectors are commercially available. The attenuation curves can be characterized as shown in figure 4. The very low

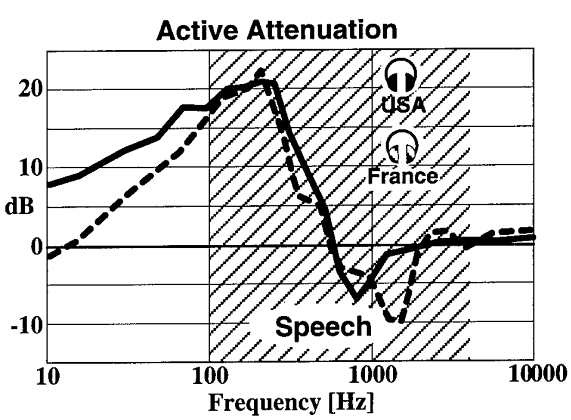


Fig. 4: Typical attenuation curves for the active part in an ANC Headset

frequencies <50Hz usually show an attenuation not exceeding 10dB. The frequency range that displays the largest attenuation is usually between 100 and 250Hz, and the maximum attenuation may be as high as 20dB. It then diminishes. Usually for frequencies higher than 800Hz no more active attenuation is shown. In the frequency range between 800 and 2000Hz all systems show an amplification of the residual noise. For frequencies higher than 2000Hz no effects are usually seen.

5. EXPERIMENTAL ANC HEARING PROTECTORS

5.1 ANC ear plug

The performances of the commercially available active hearing protectors, correspond quite well to the problems that have been shown in the introduction to this paper. They add up to 20dB of attenuation in the low frequency area of a hearing protector. The allowed exposure time for the noise shown in figure1 would now be more than 8 hours. The masking of the speech also would be reduced by the added noise reduction at low frequencies.

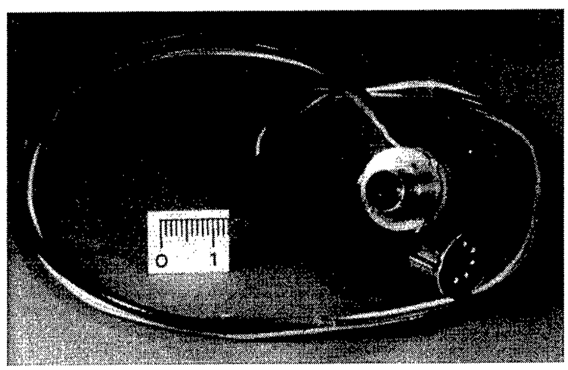


fig. 5: Experimental ANC ear plug

However, all devices show a amplification up to 10dB at frequencies around 1kHz. This amplification can be quite disturbing because it amplifies residual noise in an area, where the passive attenuation is not yet maximum and where the speech may be disturbed. As this effect is always situated just after the effective area of the active attenuation, this bandwidth has to be extended in order to shift the amplification area to a spectral part where it may not impede the speech signal. The best way to achieve this, is to reduce the volume of the device, and this leads directly from an ear muff to an ear plug. Figure 5 shows a photograph of such an experimental active ear plug that has been made by the ISL. The size of this plug is about 2x2cm. The "loudspeaker" for it is a specially designed piezo-ceramic transducer. The microphone is a standard miniature microphone. The size of this "ear-plug" is still to big to fit comfortably in the ear canal, but the active attenuation of the plug (figure 6), indicates that this is a possibility for further development of active hearing protection. The active attenuation of this ear plug shows, that it is possible to extend the usable frequency range towards higher

frequencies (~3kHz in figure 6). In this way the area

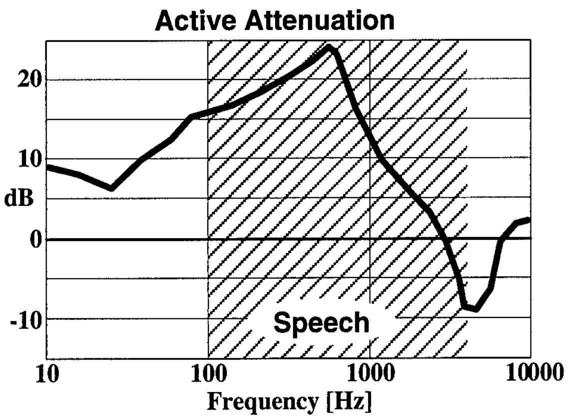


fig. 6: Active attenuation of an experimental ANC ear plug

where the residual noise is amplified is shifted to higher frequencies, where the passive attenuation is very good and where the influence on the intelligibility of speech is reduced.

5.2 digital ANC systems

5.2.1 digital feedback systems

All commercially available ANC hearing protectors have the compensation filter in the feedback loop (F in figure 3) implemented as a analog filter (fig. 7a). This type of filter is especially designed for one type of electroacoustic system, and one type of external noise. If a hearing protector must be adapted to another type of electroacoustic system or to another type of external noise, the feedback compensation filter has to be redesigned and implemented as a new piece of hard ware. If the compensation filter is implemented as a

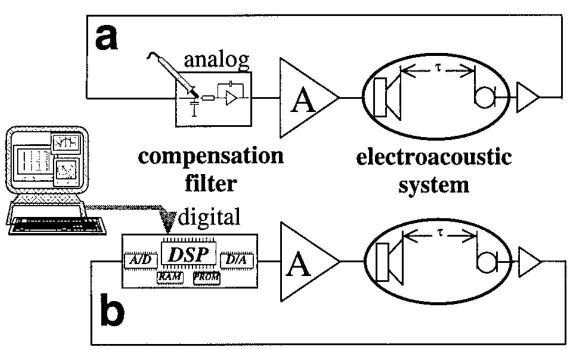


fig. 7: different possibilities for the implementation of the feedback compensation filter
a) analog
b) digital

digital filter (fig. 7b) using a DSP [5] (Digital Signal Processor) a modification of the electroacoustic or another type of external noise, only means a download of new parameters from a computer. In this type of feedback systems it is even possible to change the filtering characteristics on the fly, if the parameters for different noise types are already loaded in the memory of the DSP system.

5.2.2 digital feedforward systems

Whereas digital feedback systems mainly reproduce the analog approach of ANC by using the gain of flexibility of a DSP driven system, the feedforward approach of ANC, that has been proposed by ELLIOT [6] and PAN [7], is a very different method. This

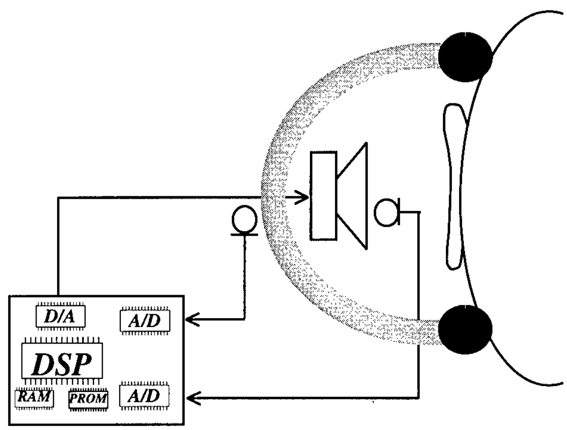


fig. 8: Principle of a digital feedforward system for a ANC Headset

method measures the noise outside the earmuff and predicts the signal at the error microphone inside the earmuff using a model of the transfer function between the two microphones. The predicted signal is then phase inverted and injected into the earmuff. Methods like that allow to adapt continuously the model of the transfer function, and to become less dependent of the fit of the protector.

5.2.3 ANC a part of the communication system

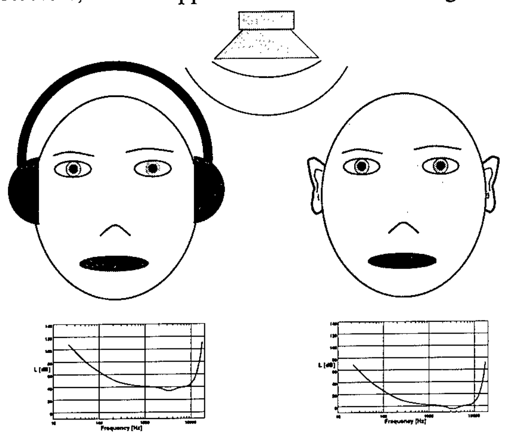
Future communication systems will be all digital. This will allow to include 3D-audio displays, voice activated switches and a lot more features that will need DSP systems. Digital ANC hearing protection in a context could be directly implemented as a part of the communication system, and not as a special device. The advantage of such a concept would be, that the optimization of the system will be for communication and protection, and not for protection only as it is for most of present systems.

6. EVALUATION PROCEDURES FOR ANC HEADSETS

6.1 Subjective method to evaluate the insertion loss

The presently normalized evaluation method for hearing protectors is the so-called "threshold method" (ISO 4869). This method (figure 9) consists in comparing the threshold in quiet of the same subject in the free sound field, with and without the hearing protector. The difference between the two threshold curves is then defined as the insertion loss (IL) of the hearing protector. The method is based on the fact, that the hearing threshold in quiet is unaltered with and without the hearing protector. Using active hearing

protectors, this supposition will not be longer true.



IL = L_{Th} protected - L_{Th} unprotected

fig. 9: Threshold method for the evaluation of hearing protectors (DIN/ISO 4869)

Figure 10 shows schematically the limitations of an active hearing protector within the area of hearing. We can see that for an important part of the frequency range, the threshold of hearing with the protector will not be limited by the curve of the threshold in quiet, but by the making effect of the electronics of the hearing protector. Using the threshold method for such

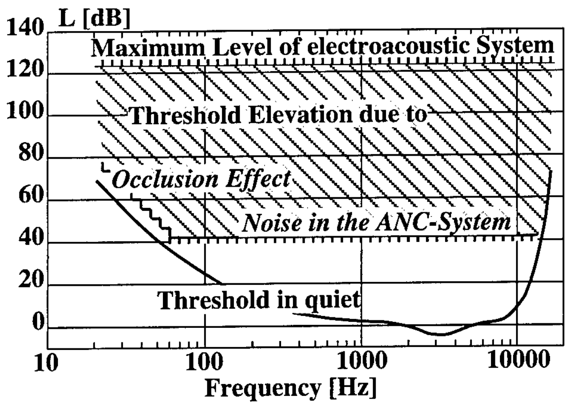


fig. 10: Schematical limits of the electro- acoustical System of an ANC hearing protector

systems would lead to an overestimation of the IL up to 20dB for some frequencies. This effect shows, that methods based on the hearing threshold of subjects, may, due to the system noise, not be adequate for the evaluation of the insertion loss.

6.2 Objective methods to evaluate the insertion loss

Objective methods are methods that measure the noise outside and under the hearing protector in order to determine the insertion loss. Figure 11 shows different methods that are in use.

6.2.1 the MIRE method

The MIRE (Microphone In Real Ear) method (fig. 11a) measures with a microphone that is positioned inside the ear canal of a subject. The insertion loss is then determined as the difference between a measurement with and without the hearing protector. This measurement has the advantage that it is done with real heads and allows to appreciate statistical distributions of a population. As the morphology of the ear canal changes between subjects, and so does the impedance of the outer ear, the two measurements (with and without protector) have to be made on the same subject. Measurements with ear plugs demand a modification of the measured ear plug. As it is not known how those modifications may influence the quality of the object to be evaluated, the MIRE method may not be the first choice.

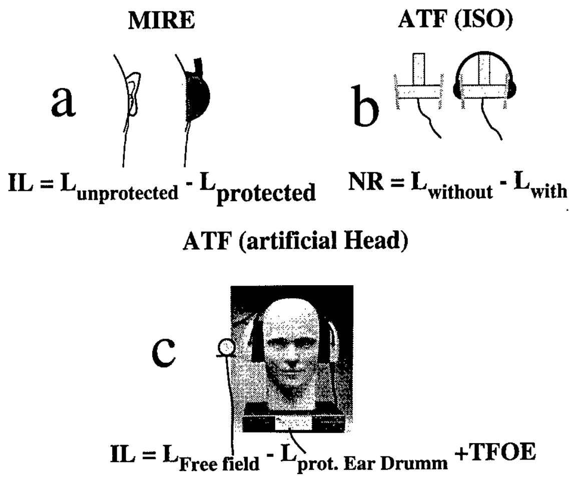


fig. 11: different objective evaluation methods for the insertion Loss (IL)
a) microphone in real ear (MIRE)
b) artificial test fixture (ATF) ISO
c) artificial head with ear simulator

6.2.2 the ISO ATF

Measurements with an ISO-type ATF (Artificial Test Fixture)(fig. 11b) are not really suitable for active hearing protections. As the quality of such protectors depends also on the impedance of outer ear, and the ISO-ATF doesn't include an ear simulator, results with this method may be questionable. Measurements with earplugs are, as no outer ear canal is present, impossible.

6.2.3 the artificial head

Artificial heads (fig. 11c) are usually equipped with outer ears (pinna + ear canal) and with ear simulators, that reproduce the impedance of the ear drum at the end of the ear canal. They give reproducable results that are close to those measured with the MIRE or Threshold method. They are also usable for the measurement of ear plugs. However the results do not show the statistical variations of a real population. The artificial heads can also be used for measurements at very high levels e.g. to show the nonlinearities of hearing protectors, where methods using human subjects cannot be used for ethical reasons.

6.3 Other measurements for the evaluation of active hearing protectors

As active hearing protectors are electronic devices, the only measurement of the IL is not enough to

characterize the quality of a device. There are at least three other parameters that should also been taken into account.

- the electrical noise that is injected into the volume of the ear cup. This noise may disturb the subjects especially when the device is used in a quiet area. It may also, in conjunction of the amplification of the feedback system (see § 4), lead to problems with speech intelligibility.
- the stability of the system. Some systems show e.g. instability if not well fitted to the head. The maximal amplification of the feedback loop (§ 4) allows to evaluate how close the system is to the stability limit.
- the behavior under high level impulse noise [8]. It is important, that the systems, when driven into a range of pressure where the electronic system overloads, behave in an acceptable way (no instability, or excessive amount of distortion).

7. Conclusion

The active hearing protection devices can be very useful for tank or aircraft crews. They should, if carefully designed, positively influence the intelligibility of the communication systems in modern weapon systems. In order to optimize the whole communication system, future generations should already include active protection and it should not (like it is today) be an independent plug-on device.

In the area of standards there is lack for the evaluation of active hearing protectors. As this protectors are electronic devices the actual evaluation procedures are not anymore adequate and/or suitable.

REFERENCES:

1. Lueg, P., "Process of silencing sound oscillation", U.S. PAtent no. 2043 416, 1936
2. Jessel, P., Angevine, O.L., "Active attenuation of a complex source", Proceedings Inter-Noise 80, 1980, pp 689-694
3. Leventhall, H.G., "Atténuateurs actifs. Développement et état du problème", Revue d'Acoustique, 59, 1981, pp 267-272
4. Franke, R., Carme, C., Vermorel, J., Dancer, A., "Protecteurs auditifs à atténuation active: étude d'orientation et de faisabilité", Rapport ISL, R-126/86, September 1986
5. Franke, R., Buck, K., Billoud, G., Sunyach, J., "Protecteur auditif miniaturisé à atténuation active utilisant un filtre numérique", Rapport ISL, R-121/89, September 1989
6. Elliot, S.J., Sutton, T.J., Rafaely, B., Johnson, M.,"Design of feedback controllers using a feedforward approach",Proceedings Active-95, 1995, pp 863-874
7. Pan, G.J., Brammer, A.J., Zera, J., "Application of adaptive feed-forward active noise control to a circumaural hearing protector",Proceedings Active-95, 1995, pp 1319-1326
8. Buck, K., Parmentier, G., "Evaluation of active hearing protectors with high level impulse noise", Proceedings of Inter-noise 96, 1996, pp 989-992

EVALUATION EN LABORATOIRE DE PROTHESES AUDITIVES DEVELOPPEES POUR L'AVIATION DE COMBAT

Médecins en Chef J.M. Clère et G. Ossard
Laboratoire de Médecine Aérospatiale
Centre d'Essais en Vol
F- 91228 Brétigny-sur-Orge Cedex.

Professeur Grateau P. (Expert oto-rhino-laryngologiste)
5 rue Manuel
F-75009 Paris

G. Reynaud
Sextant Avionique
F-33166 Saint-Médard en Jalles

RESUME Le dialogue homme-machine est amélioré par l'installation de nouveaux systèmes (viseur visuel) sur les équipements de tête (casque, masque,...). Ces systèmes alourdissent les équipements, voire les rendent dangereux sous facteur de charge. Pour alléger les équipements de tête, il a été envisagé de remplacer les écouteurs par des prothèses auditives (PA), « bouchons d'oreille » de moindre poids, équipés de transducteurs miniatures pour le transfert de messages phoniques. **METHODOLOGIE.** Les effets du port des PA lors de situations aéronautiques (accélération et altitude) ont été évalués en laboratoire. Un modèle analogique de tête humaine reproduisant les cavités aériennes (bouche, nez, sinus et oreilles) a été développé et équipé de cinq capteurs de pression et d'un accéléromètre trois axes. Quatre types de bouchons d'oreille du commerce et un bouchon personnalisé ont été successivement mis en place dans les oreilles et testés. Les effets des accélérations ont été étudiés en centrifugeuse (2 à 9 $+G_z$ établi à 0.8 G/s) et sur une rampe de siège éjectable (18 G à 300 G/s). Les effets des variations de pression ont été étudiés en chambre d'altitude (1013 à 300 hPa établi à 500 hPa/min et des variations de 300 hPa en 0.2 s). **RESULTATS.** Les accélérations n'ont pas entraîné de déplacement des PA par rapport aux structures anatomiques. Lors des variations de pression barométrique, l'étanchéité de certaines PA pourrait être à l'origine d'une différence de pression de part et d'autre du tympan entraînant un risque lésionnel élevé. **DISCUSSION.** Pour éviter cette différence de pression, il est nécessaire de ménager une communication aérienne entre les faces internes et externes des futures PA personnalisées. Cette expérimentation permet d'établir les spécifications de PA légères susceptibles de remplacer sans danger les écouteurs des casques actuels.

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

1. INTRODUCTION

Dans le cadre du dialogue Homme-Machine des pilotes d'avions de combat de nouvelle génération, il est prévu de mieux utiliser le canal auditif et le canal vocal. Il s'agit pour le pilote soit d'obtenir des informations nouvelles par le canal auditif soit de donner des ordres à la machine par l'intermédiaire de la voix. C'est ainsi que les concepts de son 3D et de commande vocale ont été développés et font l'objet d'études de validation. L'utilisation de ces concepts se heurte toutefois à l'environnement très bruyant de l'avion. Il faut donc améliorer le rapport signal-bruit en réduisant de façon passive ou active le bruit ambiant et en optimisant la transmission des signaux acoustiques. De plus, il est nécessaire de réaliser ce but en limitant le poids du casque. L'architecture du système fait appel à des prothèses auditives, installées dans le conduit auditif externe, et dotées de haut-parleurs et de micros.

Le système prototype a été réalisé par la société SEXTANT à la demande du Service Technique des Télécommunications et des Equipements Aéronautiques. Les prothèses auditives utilisées ne doivent pas présenter d'inconvénients majeurs pour les utilisateurs. En effet, au cours du vol, il existe des variations de pression cabine et de gravité. Les variations de pression pourraient être à l'origine d'algies ou de traumatisme du tympan. Les variations de gravité pourraient être à l'origine de déplacement des prothèses ce qui altérerait les communications phoniques de l'équipage avec son système de bord ou avec le contrôle. Pour s'en assurer, une étude expérimentale a été réalisée au Laboratoire de Médecine Aérospatiale (LAMAS). Pour des raisons de coût, les prothèses auditives ont été remplacées par des bouchons d'oreille.

L'expérimentation a eu donc pour objet d'évaluer le comportement de ces bouchons d'oreille dans deux conditions expérimentales :

- variations de pression simulant l'altitude cabine d'un avion

- variations d'accélérations simulant celles observées dans un avion.

De plus, cette expérimentation a été menée en prenant en compte les situations de surpression ventilatoire sous accélération ou en altitude.

2. METHODOLOGIE

Ce chapitre présente les bouchons d'oreille, le dispositif expérimental développé dans le but de cette évaluation, les moyens d'essais et de mesure, les protocoles expérimentaux.

2.1. TYPES DE BOUCHONS D'OREILLES

Pour pouvoir disposer de différents types et formes de bouchons d'oreille du commerce, dix modèles différents ont été sélectionnés. Etant donné la similitude de certains d'entre eux, cinq bouchons ont été retenus :

- BILSTOM « Quiestone », de couleur orange, ils comportent un volet d'étanchéité et sont évidés dans leur partie centrale vers l'extérieur du conduit auditif

- EAR « Form », de couleur jaune, ils sont dotés de trois volets d'étanchéité

- RACAL « Air soft », de couleur bleue, ils sont dotés de trois collets d'étanchéité enserrant une pièce cylindrique en polymère

- RACAL « dBa », de couleur orange, ils sont constitués d'une partie extérieure bombée en polymère souple dans laquelle

est placée une pièce cylindrique rigide lors de la mise en place de ces bouchons.

- Bouchons personnalisés, réalisés par des audioprothésistes. Ces bouchons sont adaptés à la forme du conduit auditif de l'utilisateur.

2.2. DISPOSITIF EXPERIMENTAL

Le dispositif expérimental est composé d'une tête artificielle réalisée par le Service de stomatologie de l'HIA Begin, Saint-Mandé, Val de Marne, France. La réalisation de cette tête artificielle a été obtenue à partir d'un moulage de crâne réel. Elle est composée d'une boite crânienne possédant ses cavités et elle est recouverte d'une peau artificielle (figure n°1). Cette tête est divisée en deux par un plan de coupe transverse et horizontal passant par les cavités aériennes de la face (cavités nasales, sinus nasaux, trompes d'Eustache, oreilles moyennes). La découpe permet une mobilité de la calotte crânienne par rapport au reste du crâne pour la mise en place des différents capteurs (décrits dans le chapitre 2.1.4.).

Cette tête est montée sur un mannequin anthropomorphique ALDERSON Hybrid II détenu par le LAMAS. Les différents bouchons sont mis en place de façon successive dans les conduits auditifs externes de la tête artificielle.

2.3. MOYEN D'ESSAIS

Les moyens d'essais prévus dans cette expérimentation sont ceux du Laboratoire de Médecine Aérospatiale (caisson d'altitude et centrifugeuse humaine) ainsi que la rampe d'éjection du Centre d'Essais en Vol.

2.3.1. Caisson d'altitude

Le caisson d'altitude permet d'effectuer les expérimentations concernant les variations de pression. Le mannequin, muni de sa tête, est installé sur un siège éjectable lui-même monté dans le caisson de 10 m^3 appelé par ailleurs caisson SAS. Le caisson de 60 m^3 est utilisé comme réserve de vide. Pour les situations de décompression rapide, la porte reliant le SAS au caisson de 60 m^3 est équipée d'une vanne à ouverture programmable. Pour les situations de décompression explosive, cette porte est remplacée par une autre porte munie d'un Rhodoïd déchirable de façon quasiment instantanée lors de sa percussion par un marteau tranchant. Le banc de régulation IN 439-5 "avionnable" est utilisé pour les situations de surpression ventilatoire en altitude.

2.3.2. Centrifugeuse

La centrifugeuse humaine est utilisée avec son siège Martin Baker MK X, dont le dossier est incliné à 20°. Le mannequin est maintenu en place par les sangles du siège. Pour les situations de surpression ventilatoire, le banc de régulation électronique L'Air Liquide est utilisé.

2.3.3. Tour d'éjection

La tour d'éjection, installée près du Laboratoire de Médecine Aérospatiale, est équipée d'un siège Martin Baker MK IV. Cette tour permet de tester des sièges éjectables ou des équipements portés par un mannequin assis sur ce siège lors de la phase canon.

2.4. METROLOGIE

Comme cela a été évoqué au cours du chapitre 2.2., la tête est munie de capteurs de pression et d'accélération.

2.4.1 Mesure de pression

La pression est mesurée au niveau du conduit auditif externe dans la zone d'air piégée entre le bouchon d'oreille et le tympan. Cette pression est appelée pression pré-tympanique. La pression est aussi mesurée au niveau de l'oreille moyenne, des fosses nasales et au niveau du masque. Lors des situations d'hypobarie ou d'accélération en centrifugeuse la pression ambiante est mesurée. Elle est appelée pression cabine.

En résumé, les pressions suivantes sont mesurées:

- Pression "pré tympanique" (un capteur de pression pour chaque oreille)

- Pression "oreille moyenne" (un capteur de pression pour chaque oreille)

- Pression "fosses nasales" (un seul capteur)

- Pression "masque"

- Pression "cabine"

Les capteurs de pression utilisés sont des capteurs ENDEVCO ± 140 hPa, capables de mesurer deux fois leur étendue de mesure sans distorsion du signal ni altération de leurs composants. Les prises de pression de "référence extérieure" situés à l'arrière des capteurs sont centralisées par des tubes souples en un point unique vers la base inférieure et postérieure du crâne. La longueur de ces tubes est similaire (même volume mort).

2.4.2. Mesure d'accélération

Un accéléromètre triaxial ENTRAN ± 50 G est fixé sur une platine métallique positionné au centre du crâne et dans sa partie supérieure.

2.4.3. Instrumentation vidéo

Pour les situations d'accélérations, obtenues en centrifugeuse ou lors d'essais sur la tour d'éjection, les capteurs de pression ne sont pas installés. Par contre, des caméras vidéo permettent de mesurer l'éventuel déplacement du bouchon par rapport aux structures anatomiques.

2.4.4. Enregistrement

Le cheminement des câbles de mesure suit celui des tubes de pression de référence des capteurs de pression. L'ensemble des données fournies par les capteurs est enregistré en mode analogique magnétique (DAT) et papier (GOULD-ES 200). Une base de temps IRIG est aussi enregistrée.

2.5.PROTOCOLE EXPERIMENTAL

Les différents bouchons d'oreille ont été évalués lors de variations de pressions barométrique et d'accélérations selon la description effectuée dans les chapitres suivants.

2.5.1. Variations de pression barométrique

2.5.1.1. Décompression lente

Le profil de décompression lente choisi est celui d'une montée d'un avion de combat récent à 12.000 m en 3 minutes (situation réaliste d'une montée de Mirage 2000 en

configuration lisse), entraînant une diminution de la pression cabine (figure n° 2). Le profil de descente est effectué à la même vitesse.

2.5.1.2. Décompression rapide ou explosive

La décompression rapide ou explosive constitue une situation exceptionnelle induite par une panne de pressurisation, une perte étanchéité du boudin de verrière voire un bris de verrière (pour le cas de la décompression explosive). Il en résulte une brusque évolution de la pression cabine vers la pression barométrique environnante. Le délai qui s'écoule entre le passage de la pression cabine vers une pression barométrique est variable selon l'origine de cette perte de pressurisation.

La situation expérimentale de décompression rapide comporte une variation de pression (entre la pression cabine et la pression barométrique) de 300 hPa. Cette situation est retrouvée lorsqu'un avion volant à 5600 m a une décompression cabine. La pression de la cabine est alors de 800 hPa et la pression barométrique environnante est de 500 hPa. pour cette expérimentation, il a été retenu cette situation de décompression avec une délai de 1 à 10 secondes.

La situation expérimentale de décompression explosive comporte une variation de pression d'environ 300 hPa avec un délai inférieur au centième de seconde.

2.5.1.3. Surpression ventilatoire

La mise en oeuvre de la surpression ventilatoire pour protéger l'équipage du risque hypoxique à haute altitude dépend uniquement de la valeur de la pression barométrique.

Le cas de la mise en oeuvre de la surpression ventilatoire est étudiée avec des valeurs de surpression ventilatoire de 100 hPa. Les délais de décompression de la cabine sont inférieurs au dixième de seconde. La surpression ventilatoire est obtenue avec l'ensemble de régulation IN-439-5 "Avionnable".

2.5.2. Variations d'accélérations

2.5.2.1. Accélérations $+G_Z$ de longue durée

Pour cette phase expérimentale, il s'agit de s'assurer de la bonne tenue mécanique du bouchon dans le conduit auditif externe. L'absence de mouvement du bouchon devrait alors permettre une conservation de l'étanchéité dans deux situations expérimentales différentes:

- accélération seule

- accélération avec utilisation de la surpression ventilatoire.

Le profil d'accélération est de 1 à 9 $+G_Z$ en montée graduelle (1G/s).

Le maintien du bouchon dans le conduit auditif externe est vérifié par les caméras vidéo à haute vitesse qui permet de repérer l'absence de déplacement du bouchon par l'intermédiaire de repères visuels "bouchons" et "structures anatomiques".

2.5.2.2. Accélérations de longue durée avec surpression ventilatoire

Lors de l'utilisation de la surpression ventilatoire, le profil d'accélération est similaire. La loi de surpression ventilatoire est de 18 hPa/G avec début de mise en pression à 4 $+G_Z$ et valeur maximale de cette surpression à 9 $+G_Z$ (90 hPa).

Cette étude comporte aussi l'étude du maintien du bouchon lors des phases de décélération et de diminution de pression ventilatoire.

2.5.2.3. Accélérations de courte durée (cas de l'éjection)

Il s'agit de s'assurer de l'absence de risque traumatique du bouchon sur les structures anatomiques qui l'environnent. Elle consiste à simuler sur la tour de siège éjectable, le départ du siège (phase canon). La tête montée sur le mannequin ALDERSON Hybrid II est soumise au départ d'un siège Martin Baker MK IV. Cette tête est équipée des différents bouchons d'oreille.

3. RESULTATS

3.1. VARIATIONS DE PRESSIONS BAROMETRIQUES

3.1.1. Décompression lente

Les mesures effectuées avec les quatre types de bouchons ont montré qu'il existait, une surpression de 2.5 hPa au niveau de la zone pré-tympanique lors de l'arrivée au pallier à 12 000 m. Cette surpression est similaire quelque le modèle de bouchon utilisé (figure 3).

Cette absence de surpression traduit une absence d'étanchéité du bouchon. Ces fuites font reculer le risque d'algies ou de barotraumatisme de l'oreille.

3.1.2. Décompression rapide

Lors des décompression rapides, le régime de pression au niveau de la zone pré-tympanique est différent d'un bouchon à l'autre :

- bouchons BILSTOM « Quiestone » et EAR « Form », aucune surpression notable n'est observée

- bouchons RACAL « Air soft », surpression pré-tympanique dont les valeurs crêtes sont de 24 et 48 hPa selon que la variation de pression barométrique est établie en 10 ou 1 secondes, (figure n° 4)

- bouchons RACAL « dBa », surpression pré-tympanique dont les valeurs crêtes sont de 7 et 36 hPa selon que la variation de pression barométrique est établie en 10 ou 1 secondes,

- Bouchons personnalisés, surpression pré-tympanique dont les valeurs crêtes sont de 20 et 70 hPa selon que la variation de pression barométrique est établie en 10 ou 1 secondes,

Ces valeurs montrent que les bouchons présentent une certaine étanchéité lors de variations rapide de pression.

3.1.3. Décompression explosive

Lors des décompression explosives, le régime de pression au niveau de la zone pré-tympanique est différent d'un bouchon à l'autre mais aussi en fonction de la variation d'altitude. Si une variation de pression barométrique de 300 hPa est produite par l'explosion du Rhodoïd séparant les deux chambres d'altitude, deux situations expérimentales ont été obtenues. En effet, il a été reproduit une variation de la pression cabine de 300 hPa mais de façon telle que la pression finale, qui est alors l'équivalent de la pression barométrique, soit de 500 ou de 105 hPa. Dans la première circonstance la surpression ventilatoire altimétrique n'est pas présente alors que dans la seconde, elle est fonctionnelle.

- bouchons BILSTOM « Quiestone », surpression prétympanique dont les valeurs

crêtes sont de 115 et 145 hPa selon que la variation de pression finale est de 500 ou 105 hPa

- bouchons EAR « Form », aucune surpression prétympanique notable n'est observée quelque soit l'altitude pression finale

- bouchons RACAL « Air soft », aucune surpression prétympanique notable n'est observée lorsque la pression cabine est de 500 hPa. Par contre, une pression prétympanique crête de 260 hPa est observée lorsque la pression finale est de 105 hPa.

- bouchons RACAL « dBa », aucune surpression prétympanique notable n'est observée lorsque la pression cabine est de 500 hPa. Par contre, une pression prétympanique crête de 82 hPa (figure n°5)est observée lorsque la pression finale est de 105 hPa.

- bouchons personnalisés, les surpressions pré-tympaniques sont de 43 et 130 hPa lorsque les pressions cabines finales sont de 500 et 105 hPa.

3.2. ACCELERATIONS

3.2.1. Variations d'accélération de longue durée

Les accélérations de longue durée sont reproduites par la centrifugeuse. La variation d'accélération et la valeur maximale d'accélération étaient respectivement de 1 G/s et de 9 +Gz.

Au cours de ces montées en accélération, il n'a pas été observé de déplacement des bouchons par rapport au canal auditif externe.

L'utilisation de la surpression ventilatoire n'a pas mis en évidence de régime de pression dangereux.

3.2.2. Accélérations de courte durée (cas de l'éjection)

Les accélérations reproduites par l'éjection ont été au début des essais à l'origine « d'une fracture occipitale » de la tête artificielle. En effet, l'interface mécanique entre le cou du mannequin est venu s'enfoncer dans la résine plus molle constituant le crâne de cette tête.

Après une nouvelle récidive, les parties molles de la tête artificielle, comportant l'oreille avec son conduit auditif externe, ont été récupérées et recollées sur la tête du mannequin Alderson Hybrid II.

Les essais ont montré qu'en fonction des bouchons, il pouvait apparaître un léger déplacement de ceux-ci pour certains d'entre eux. Ce déplacement est resté limité pour les bouchons achetés dans le commerce. En revanche, il n'a pas été observé de déplacement des bouchons personnalisés.

Ces essais n'étaient toutefois pas totalement représentatif des conditions réelles du port des bouchons puisque, pour pouvoir observer l'éventuel déplacement, le mannequin n'était pas équipé de casque. Or ce casque peut avoir un rôle de contention.

4.DISCUSSION-CONCLUSION

Cette expérimentation a montré que le port de bouchons d'oreille sans dispositif particulier pouvait entraîner des algies ou des barotraumatismes de la sphère O.R.L. lors de situations exceptionnelles représentées par les décompression rapides ou explosives. En effet, il apparaît des valeurs de surpression au niveau de la zone prétympanique pouvant atteindre plusieurs centaines d'hectoPascal. Il serait donc nécessaire de réaliser, pour chaque bouchon, un évent permettant un rétablissement plus rapide de la pression dans la région prétympanique.

En revanche, lors des accélérations de longue ou de courte durée, il ne semble par exister de risque lésionnel par déplacement des bouchons.

Des études complémentaires doivent être menées à l'aide d'expérimentation humaine pour s'assurer de l'innocuité de ce dispositif.

IMPAIRED NOISE-ATTENUATION OF AIRCREW HELMETS AND HEADSETS FOR COCKPIT PERSONNEL WHO WEAR GLASSES

D.-M. Rose*, H. Welsch**, H. Pongratz**, J. Konietzko*

Institute of Occupational-, Social- and Environmental Medicine*,
Johannes Gutenberg-University of Mainz
Obere Zahlbacher Str. 67, D-55131 Mainz, Germany
and German Air Force Institute of Aerospace Medicine, Fürstenfeldbruck, Germany**

SUMMARY

Goggles significantly reduce the noise attenuation provided by hearing protection. The alteration of noise attenuation in 3 different helmets and one headset with 4 different spectacles was the object of this investigation. Sound pressure levels were measured inside the auditory canals of 11 candidates who were exposed to pink noise of 104 dB(lin) SPL with and without wearing the different types of spectacles and helmets. The mean noise attenuation of the headset and the helmet No 1 with separate ear-cuffs (SPH-4) was reduced in the mean up to 6 dB by glasses with thick horn-rimmed frames and less by glasses with thin metal frames. Helmets No 2 and No 3 (HGU-55 and an integrated helmet) provided only poor noise protection, but there was no further reduction of noise attenuation by wearing glasses. Headsets and helmets with separate ear-cuffs provided good noise protection. The reduction of noise attenuation with spectacles is significant depending on the thickness of the ear-piece. Thick hornrims could potentially increase the risk of hearing impairment. If noise attenuation values are already poor (integrated helmet) glasses will not change the values much. To avoid hearing damage, only spectacles with thin frames should be worn by aircrews. In addition the visual field will also be enlarged.

1 INTRODUCTION

Investigations in occupational medicine have shown that goggles with large plastic or horn-rimmed frames significantly reduce the noise attenuation provided by personal hearing protection devices. No investigations were conducted for the influence of spectacles on the noise attenuation of personal hearing protection. The object of this study design was to test the changes in noise protection provided by 3 helmets and one headset without glasses in comparison with the results affected by wearing different spectacle designs

2 METHODS

The influence of 4 different spectacles on the alteration of noise attenuation of 3 helmets and one headset was investigated.

2.1 Spectacles

Spectacle design No 1 is approved for aircrew use in the Federal Armed Forces. The earpiece is of thin steel, with a classic curved end. Spectacle design No 2 is approved for aircrew use in the US Air Force, and was tested for the use in the Federal Armed Forces. The difference between spectacle design No 1 and No 2 is an erect ear piece in No 2. Spectacle design No 3 has a normal horn-rimmed frame. The horn-rimmed earpiece is a little bit thicker than the steel pieces of spectacles No 1 and 2. Spectacle design No 4 is a modern, fashionable horn-rimmed frame with thick earpieces as is worn for example by private pilots.

2.2 Helmets and headset

Helmet No 1 is a helicopter helmet (SPH 4). The helmet has separate ear cuffs to

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

best fit the pilot's ears. Helmet No 2 is a fighter aircraft crew helmet (HGU 55). The ear-shell is integrated into the helmet and therefore there exists a great problem in sufficient noise attenuation even without glasses. Even though pilot helmets were individually fitted, we found in another study that only 6 of 20 helmets provided the pilot with a fair to good noise-attenuation. Helmet No 3 (integrated helmet) is a modified motorcyclist helmet, which could be an example for a subsystem of a future whole-body pilot suit. The ear-shell also is integrated into the helmet. The helmet is conceived more as protection and has no special noise-attenuation concept. The circumaureal headset (Peltor) is widely used by transport crews in the Federal Armed Forces and in general aviation. The headband pressure provides a good seal for the earcup cushion.

2.3 Experimental design

The object of this study design was to test the changes in noise protection provided by the three helmets and the headset without glasses in comparison with the results affected by wearing the different spectacle designs. The tests were conducted in a room providing a diffuse acoustic field at noise levels above 200 Hz with 11 volunteer candidates who were exposed to pink noise of 104 dB SPL. The acoustic irradiation generated for the tests was pink noise (constant noise intensity throughout the entire frequency range) generated by a noise generator and amplifier (Ralph E. Behr Type ESRG-50 N) and filtered by a spectrum generator (Brüel & Kjær Type 5612) and supplied to the acoustic irradiation room via four sets of loudspeakers (Ralph E. Behr Type LK 50 T). The maximum noise level provided in this fashion amounted to 110 dB.

The noise attenuation was measured with an objective method. A miniature electret microphone was glued to a brass plate, fitted to an ear plug (Model Selektone K) and placed in the candidate's outer auditory canal.

The changes in noise attenuation using spectacles resulted from the differences between the measurements with and without hearing protection and with and without glasses. Using variable measurements the correct position of the ear plug in the outer auditory canal was determined.

The statistical evaluation between the values with and without hearing protection and with and without glass frames was done by variance analysis.

3 RESULTS

Helmet No 1 was the helicopter helmet (SPH 4). The helmet has separate ear cuffs to best fit the pilot's ears. Depending on the thickness of the ear-piece noise-attenuation was significantly reduced. In the worst case there was a difference of 12 dB between the noise attenuation without glasses and the thick horn-rimmed frame. Up to 1000 Hz with spectacle No 4 (the thick horn-rimmed frame) there is no significant noise-attenuation (Tab. 1, Fig. 1).

Helmet No 2 was the fighter aircraft crew helmet (HGU 55). The ear-shell is integrated into the helmet. Results showed only little effect on noise-attenuation by spectacles. Only in higher frequencies one has found the same characteristics of noise-attenuation differences as in helmet No 1 but with a much smaller spread. The reason may be the big pad of foam rubber, which fitted tightly to the frame's ear-pieces (Tab. 1, Fig. 1).

Helmet No 3 was a modified motorcyclist helmet. The ear-shell also is integrated into the helmet. The ear-shells did not fit well and the results showed a very poor noise-attenuation. In lower frequencies there was an enhancement of the ambient noise inside the helmet by air-vibration. No differences were seen between the results for persons wearing glasses or not. In single cases the results were even better with spectacles than without, probably by filling airspaces

between the subject's ear and the ear-shell (Tab. 1, Fig. 1).

The headband pressure of the circumaureal headset provided a good seal for the earcup cushion. Therefore the optimal noise-attenuation was better with the headset than the helmets. There was also a pad of foam rubber which fitted tightly to the frame's ear-pieces but great differences between different spectacle designs occurred. In the worst case there was a significant 16 dB reduction in noise-attenuation by frame No 4 (Tab. 1, Fig. 1).

4 DISCUSSION

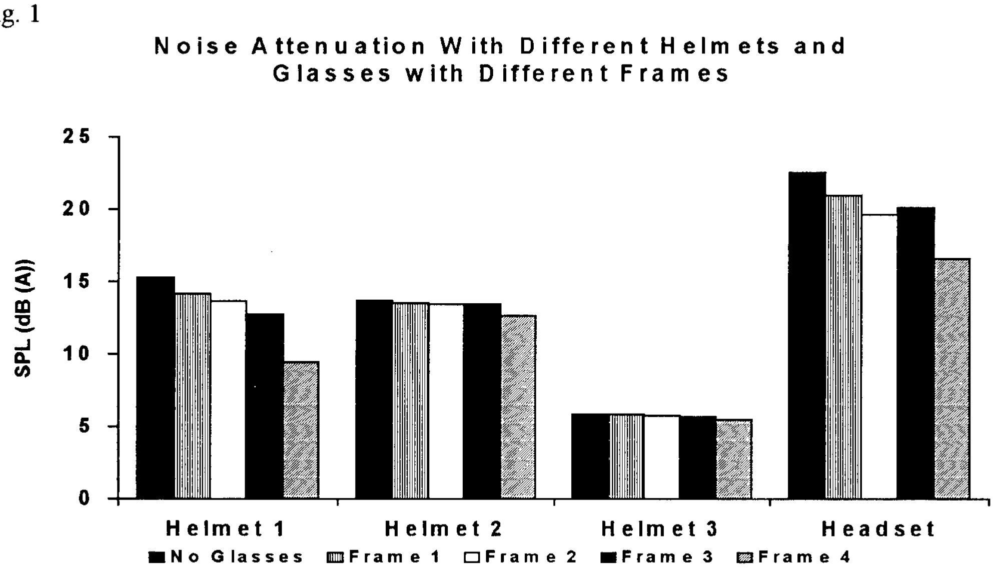
The headset and the helmet with separate ear-shells fit best to the ear and provide the best noise attenuation. The reduction of noise attenuation with spectacles is significantly depending on the thickness of the ear-pieces. Helmets with integrated ear-shells are not substantial affected in noise-attenuation by spectacles. But a noise reduction of 15 dB versus the ambient noise means only a 5 dB noise reduction if the essential noise to ratio spread is added for the radio communication. Therefore the time for a crew exposed to aircraft noise during operations without the risk of hearing-damage is small.

In addition even small differences in noise attenuation have great effects on the speech intelligibility. In a further study with the headset and spectacle design No 1 the difference in noise attenuation was only 1,5 dB in mean by wearing glasses, but the error rate for wearer of glasses rose 4 % for monosyllabic words.
In another study we found that wearers of glasses with the jet helmet had a slightly higher error rate in speech intelligibility tests than persons with a mild or moderate hearing loss without wearing glasses.
These results may have an important impact on flight safety, partcularly if a higher reduction in noise attenuation by thicker ear-pieces in frames occurs, perhaps in combination with a mild or higher hearing-damage in older pilots.

Therefore we propose:

- An improvement in the noise attenuation in helmets if necessary with electronic aids like Active Noise Reduction systems.
- Wearers of glasses should only wear glasses with a thin frame, not only to avoid hearing damage but also to enlarge the visual field.
- For presbyopic military pilots the use of contact lenses should be possible or should even be recommended.

Fig. 1



Tab. 1

Noise Attenuation	**Without Glasses**	**Spectacles No.1**	**Spectacles No.2**	**Spectacles No.3**	**Spectacles No.4**
Headset					
mean dB(A)	22,61	20,98	19,66	20,14	16,60
std	2,93	2,95	3,17	3,01	4,07
min	15,40	14,80	14,60	15,30	10,10
max	26,10	25,40	24,30	24,70	24,90
mean 1000 Hz	29,20	27,40	25,70	26,44	22,56
std	5,14	4,92	4,74	4,28	5,37
min	17,70	17,10	17,10	17,80	14,10
max	34,90	34,60	34,60	31,70	32,70
significance					+
Helmet No. 1					
mean dB(A)	15,29	14,18	13,65	12,79	9,46
std	2,08	2,05	2,25	2,10	3,43
min	11,60	11,20	10,90	10,20	4,90
max	18,30	17,30	17,00	16,70	14,50
mean 1000 Hz	14,24	13,77	13,47	13,14	10,17
std	2,03	2,21	2,53	2,47	3,58
min	11,60	10,40	9,40	9,20	5,20
max	18,70	18,60	18,90	18,70	16,80
significance					+,*
Helmet No. 2					
mean dB(A)	13,71	13,54	13,41	13,45	12,70
std	2,29	2,31	2,36	2,27	2,47
min	9,50	9,90	9,50	9,60	8,90
max	17,00	16,80	16,80	16,60	16,40
mean 1000 Hz	17,52	17,39	17,42	17,56	16,51
std	2,04	1,99	2,12	2,03	2,37
min	13,40	13,80	13,20	13,50	12,30
max	20,50	20,40	20,30	20,40	19,50
significance					
Helmet No. 3					
mean dB(A)	5,84	5,80	5,71	5,63	5,45
std	1,60	1,51	1,48	1,48	1,47
min	1,80	2,10	2,10	2,10	1,90
max	7,60	7,50	7,50	7,50	7,20
mean 1000 Hz	6,35	6,26	6,06	6,08	5,80
std	1,98	1,87	1,83	1,80	1,83
min	1,20	1,40	1,40	1,40	1,20
max	8,60	8,30	8,20	8,20	8,10
significance					

mean= mean noise attenuation at 1000 Hz or over all frequencies in dB (A), std= standard deviation, min = minimum noise attenuation for n = 11, max = maximum noise attenuation for n = 11, significance= $p< 0,05$,+ = compared to the values without glasses, * = compared to spectacles No. 1

THE APPLICATION OF A PROPRIETARY SOUND-ATTENUATING TECHNOLOGY TO PASSIVE CIRCUMAURAL HEARING PROTECTOR DESIGN

G. B. Thomas
D. W. Maxwell
P.R. Van Dyke
Naval Aerospace Medical Research Laboratory
51 Hovey Road
Pensacola, FL 32508-1046, USA

SUMMARY

The United States Navy recently patented (U.S. Patent #5,400, 296) a composite technology that significantly improves a base material's ability to attenuate acoustical energy, particularly low-frequency acoustical energy. Given our success in applying the technology to components used by the transportation industry, we decided to investigate the feasibility of applying the technology to materials useful in the fabrication of circumaural hearing protectors.

The proprietary technology is based on maximizing characteristic acoustic impedance differences between the constituents of the composite material. Because each base material used in the construction of the various components comprising a circumaural earcup assembly generally possesses a different inherent characteristic acoustic impedance, specific composite formulas had to be derived for each component material. That is, empirically derived formulas were required for the earcup shell material (i.e., epoxy resin), the ear seal material (i.e., silicone rubber), the ear seal filler (i.e., silicone gel), and requisite adhesives (i.e., silicone sealers).

Hearing protector components were fabricated, then modified if necessary, based on results from flat plate coupler tests. Concentrating on noise fequencies below 125 Hz, we were able to fabricate earcup components that were generally superior in noise attenuation to those currently in standard use. In some instances, performance on the flat plate coupler yielded attenuation gains (relative to standard issue hearing protectors) of about 20 dB (at 31.5 Hz, for example). Gains on human models below 125 Hz are in the 9-15 dB range. The weak link in the earcup assembly remains the traditionally problematical ear seal (and the inverse relationship between noise-attenuation effectiveness and user comfort and acceptance). New materials and designs are being investigated to optimize this component.

BACKGROUND

This paper is the result of research conducted under the project, "Enhanced Hearing Protection for High-Noise Environments," funded by the Naval Medical Research and Development Command and the Naval Air Warfare Center. The purpose of the project is to develop a new type of low-cost hearing protector for use in very high noise environments. We are particularly interested in improving the attenuation of low-frequency noise because our initial target population is helicopter pilots, and low-frequency noise (below about 125 Hz) is a particular problem in helicopters.

The project began several years ago with a novel earcup design that required the use of gasket material that was especially effective at frequencies below 125 Hz. In order to retain the effectiveness of the design, we calculated that the required gasket material should attenuate low frequencies by about 25-30 dB. Unfortunately, after testing dozens of commercially available materials, we found that, while these materials were generally very good at frequencies above 500 Hz, the level of attenuation we sought at the lower frequencies was unattainable in the configuration we required. Consequently, we decided to begin fabricating gasket materials in our own laboratory.

Our first attempt at improving the sound-attenuating characteristics of a base material followed a traditional rule of thumb: "Add mass to improve attenuation." We therefore loaded a polyurethane base material with powdered lead in varying proportions. The middle curve on Figure 1 represents a polyurethane/50% powdered lead composite that was successful in meeting our attenuation criteria. (The top curve is the mean of the conventional materials tested earlier.) Although the lead-loaded polyurethane succeeded in attenuating low-frequency acoustical energy, it also succeeded in tripling the weight of the gasket (relative

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

to the commercially available materials). This weight increase was unacceptable so we began exploring other possibilities.

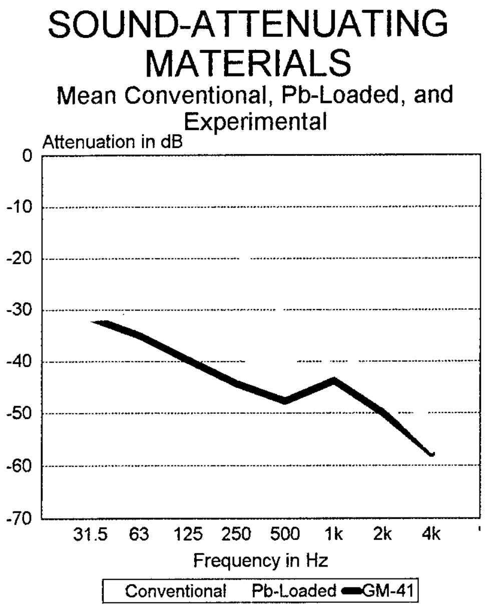


Figure 1. Sound-attenuating materials.

INITIAL RESEARCH

It occurred to us that one type of material construction, that is, the lamination of dissimilar materials, often showed improved low-frequency attenuation when compared to homogeneous materials. This is a technique that is used, for example, in the construction industry where the "floating floor" and "double wall with air gaps" techniques are popular. At least one of the principles at work in this improvement is the inefficient transmission of energy from one material to a second, dissimilar material. In other words, energy is optimally transferred from one material to a second material when the two materials are identical, and there is no intervening medium. One quality that the two materials share, and that is of particular interest to us in this application, is "characteristic acoustic impedance" (defined as the product of the mass of a material times the speed of sound through that material). It was a logical progression to the hypothesis that materials chosen on the basis of their dissimilar characteristic acoustic impedances would result in an inefficient transfer of energy, and thus, improved energy attenuation. We knew that the principle operated in layered structures or laminates; the question was, "Would we see a similar effect using very small dissimilar particles mixed into a base material?"

The bottom curve on Figure 1 represents our 41st attempt at answering the question. As can be seen, the bottom curve, labeled GM-41 (for "gasket material 41) provided the attenuation we sought and at a weight similar to the commercially available sound-attenuating materials. GM-41 is a very precise formulation of high- and low-acoustic-impedance particles in a polyurethane substrate. Following additional work and refinement, this technology was awarded United States Patent #5,400, 296. The technology is presently being used in the transportation industry.

This technology, however, is not without its qualifications. First of all, because the technology is still in its developmental stages, we continue to learn about all of the relevant variables that impact its success or failure in a given application. Second, the formulas derived for various base materials are different; that is, each base material has, so far, required a different filler material formulation for optimal attenuation. Third, the derived formulas are extremely specific; in other words, the technology is not particularly robust. Fourth, it appears that the technology can be applied to a fairly wide range of base materials. So far, we have applied it to epoxy resins, polyurethanes, silicone rubbers and gels, and carbon-based rubbers, and we are presently working on thermoplastics. Finally, virutally all of our research has been centered on airborne sound (as opposed to structure- or water-borne sound) and small surface area applications. Work has recently begun on large surface areas (i.e., sheets) and vibration applications.

APPLICATION TO HEARING PROTECTORS

Figure 2 illustrates the components of the earcup to which we have attempted to apply the proprietary technology. There are four basic components in the earcup: the earcup itself, a multi-channeled circumaural ear seal or ear cushion, a low-durometer gel within the ear seal, and a gasket interfacing the earcup and ear seal. In addition, a silicone adhesive was optimized and used to affix some of the components.

The overall size and architecture of the experimental hearing protector followed that of the HGU-84/P earcup that is currently used in the MH-54 series of

helicopters and is designed to directly replace that standard-issue hearing protector.

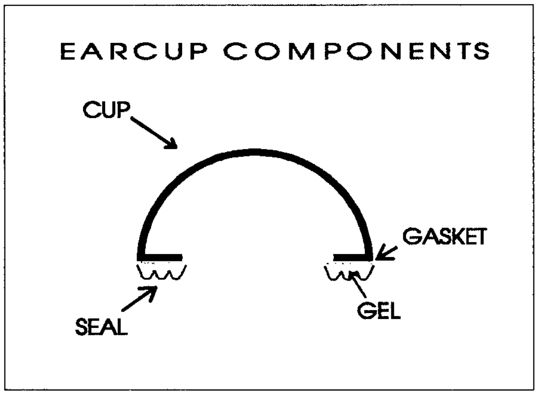


Figure 2. Earcup components.

Figure 3 shows the attenuation of a standard-issue, helmet-mounted earcup (top curve) and that of the experimental earcup (bottom curve). This is the earcup material only and not the ear seal, gel, etc. The base material of the experimental earcup is an epoxy resin to which the impedance mismatching technology had been applied. Because 40 dB of attenuation is the practical limit for hearing protectors (that is, beyond about 40 dB, bone conduction begins contributing to the noise dose), we believed that the attenuation of the experimental earcup was more than sufficient.

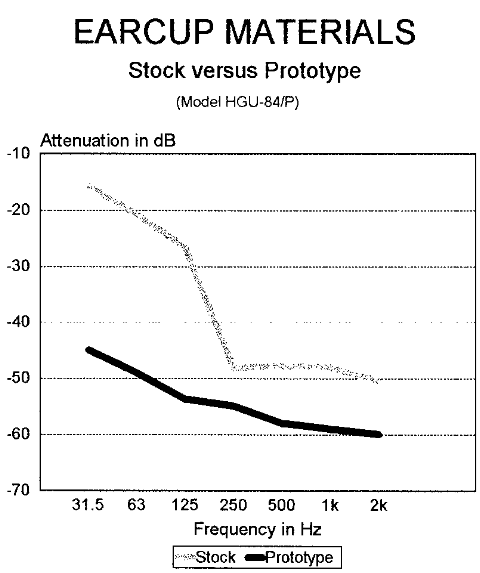


Figure 3. Earcup materials.

The ear seal or ear cushion is constructed of a silicone rubber and is configured in a series of hollow, concentric rings. The hollow, concentric rings permit the inclusion of a soft, silicone gel in the ear seal, and they also take advantage of a "mass-air-mass" architecture to increase the probability of energy loss as the noise traverses the ear seal to the hearing protector's interior. Figure 4 illustrates the effect of the technology on the silicone rubber used for the ear seal and the interface gasket. Note that this is a comparison graph of the stock silicone rubber (upper curve) with its optimized variation (lower curve) and does not reflect a comparison with the standard-issue ear seal.

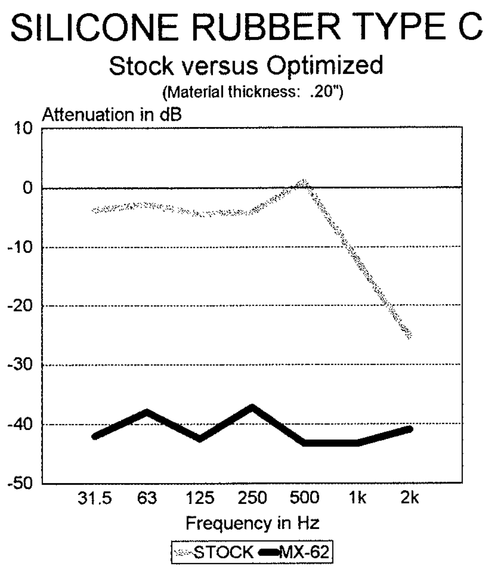


Figure 4. Silicone rubber Type C

Figure 5 illustrates the effect of applying the technology to the silicone gel ear seal filler. This material filled the hollow channels of the multi-channel ear seal and provided some measure of increased comfort. Without treatment, the stock gel was relatively acoustically transparent at the lower frequencies.

In addition to the aforementioned components, a silicone adhesive was optimized for use in assembling the components of the prototype. This material optimization of all of the prinicipal components of the hearing protector stemmed during development from

our validation of the assumption that a hearing protector is only as good as its weakest link. A great

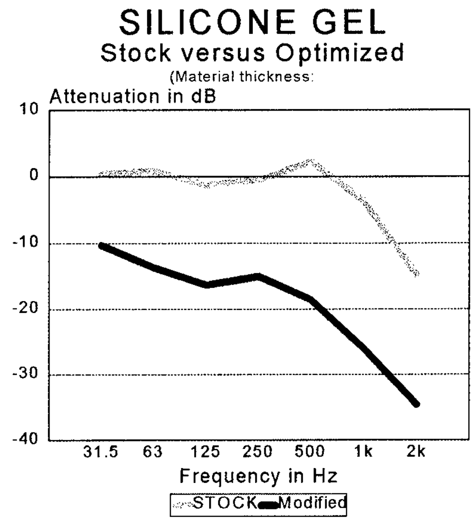


Figure 5. Silicone gel.

deal of development time was spent identifying and addressing the "currently weakest" component, then moving on to attempt to correct the next weakest.

RESULTS

Figure 6 shows the relative attenuation values of the completely assembled stock (top curve) and prototype (bottom curve) hearing protectors as measured on a laboratory flat-plate coupler. A flat-plate coupler is typically a flat, smooth metal plate into which a measuring microphone has been embedded. Measurements are taken in a noise field with the microphone uncovered and then covered by the hearing protector. Differences between the two measurements provide the attenuation values illustrated in this figure. The flat-plate coupler used in this study is of a custom design and utilizes an eight-microphone array, analog signal generation,and digital signal analysis; a bank of 34 speakers and supporting electronics provides an overall noise field of 120 dB (SPL).

As can be noted in Figure 6, the prototype hearing protector provided approximately 6-16 dB more attenuation at frequencies below 500 Hz. This corresponds to an approximate 100 - 200% improvement over the stock hearing protector. It appears to be increasingly effective at the lower frequencies. The peak at 63 Hz is apparently an artifact caused by 60 Hz line noise in our measurement equipment.

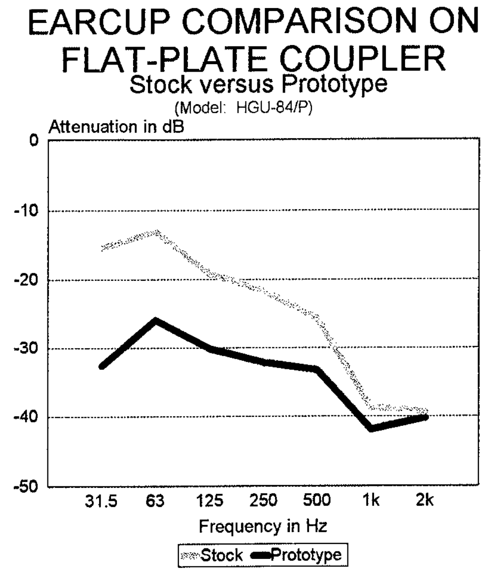


Figure 6. Earcup comparison on flat-plate coupler.

Obtaining good results on a flat-plate coupler is promising but does not always guarantee equally good results when tested on human subjects. We have had several designs that were actually superior on the flat-plate coupler to the experimental earcup described here but provided disappointing results when tested on a human model. Variables such as earcup sealing, ear seal compliance, comfort, etc. are all important in the ultimate success of a hearing protector.

Figure 7 illustrates the attenuation of the stock (upper curve) and experimental prototype (lower curve) hearing protectors when tested on the human model in compliance with ANSI Standard S12.6-1984, Method for the Measurement of the Real-Ear Attenuation of Hearing Protectors. Please note that this standard provides only for the measurement of frequencies down to 125 Hz. The data points at 63 Hz were the lowest we could obtain in our test booth and remain within the constraints of the standard. The 31.5 Hz point was derived through extrapolation but is consistent with flat-plate predictions.

The data show a 50 - 250% improvement over the standard-issue hearing protector at the lower frequencies. The peak at 250 Hz is at least partially due to one subject's unusual contribution to the data.

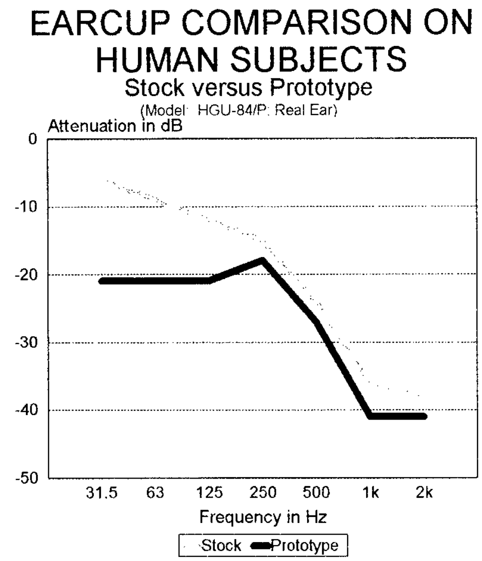


Figure 7. Earcup comparison on human subjects.

CONCLUSION

Our attempt to apply the proprietary technology to circumaural hearing protectors has been successful to a degree. The data from our early prototypeare promising, but should be able to be improved with further research. We continue to evaluate candidate materials and continue to strive to strike a balance between ear seal comfort and effectiveness; several new designs are under development.

The technology itself is still in its infancy with much work remaining. For example, there are literally hundreds of potentially useful high- and low-impedance filler materials; we have investigated less than a dozen. Applying the technology to large surface areas presents some unique problems, as does developing a spray-on version for retrofitting existing structures. Research in all of the aforementioned areas will continue in the foreseeable future.

ACKNOWLEDGEMENT

The authors would like to acknowledge the contributions of Mr. Bruce Guy and his technical staff at the Mold-Ex Rubber Co., Milton, Florida. They fabricated many of the test samples and provided expert advice on silicone rubber chemistry.

THE COMMUNICATIONS EARPLUG: A LOGICAL CHOICE FOR VOICE COMMUNICATIONS IN AIRCRAFT

Ben T. Mozo, B.S.
U.S. Army Aeromedical Research Laboratory
Fort Rucker, AL 36362-0577, USA

John E. Ribera, Ph.D.
Brooke Army Medical Center
San Antonio, TX 78234-6200, USA

SUMMARY

The U.S. Army aviator works in high levels of noise and routinely faces the challenge of effective voice communication. Existing aviator helmets, while adequate in providing hearing protection, do not provide the signal-to-noise ratio necessary to optimize in-flight voice communications. The Communications Earplug (CEP) is a small device worn by the aviator and provides significant improvements in hearing protection and communication performance. The CEP uses a miniature earphone transducer adapted to a replaceable foam earplug. Attenuation characteristics of the CEP are similar to those of other insert hearing protective devices and provide adequate protection in U.S. Army noise environments. Additional protection results when the CEP is worn with the aviator's helmet. The CEP is comfortable over a period of several hours and, in its current configuration, is considered highly acceptable by seasoned aviators and crewmembers. The CEP is easier to insert and seat in the outer ear canal than other insert protectors available through military channels. Speech intelligibility in simulated helicopter noise is significantly enhanced when using the CEP when compared to the standard SPH-4 and HGU-56/P aviator's helmets. CEP and active noise reduction (ANR) results are comparable in terms of speech intelligibility. However, there are several differences that should be considered before deciding which is the system of choice. The technology developed for CEP has wide-ranging application in the military and can easily be adapted to communication needs in the civilian community. The CEP is an inexpensive device that can enhance air and ground crewmember voice communications in the operational environment, and should be positively considered for inclusion into all aircraft and vehicular communication helmets as a battlefield multiplier for the 21st century.

1 Introduction

Noise levels found in military helicopters exceed noise exposure limits required by U.S. DOD Instruction 6055.12, "Department of Defense Hearing Conservation Program." [1] Noise levels in helicopters with higher load capacities such as the CH-47 and H-53 are extremely intense and sometimes exceed the helmet's protective capabilities. Figure 1 shows a distribution of noise levels found in U.S. Army aviation, along with estimates of noise exposure for crewmen wearing the standard protectors. Figure 2 shows the same distribution in cumulative percent for estimating the overall protection for the user population. The data show protection is adequate in all but the top 15 percent of the noise conditions while wearing the SPH-4 or HGU-56/P and in 99 percent of the cases while wearing the yellow foam earplug. Combination protection, earplugs in addition to the helmet, is a technique commonly used to provide additional hearing protection, but this technique generally decreases the aviator's ability to communicate.

The U.S. Army Aeromedical Research Laboratory (USAARL) is investigating two techniques which may be used to reduce noise exposure and improve communications. One technique, active noise reduction (ANR), uses electronic circuitry to manipulate and reduce the noise found inside the earcup. The other technique, CEP, relies on passive sound attenuation of the earplug in combination with the earcup to achieve the required noise reduction. Both systems show significant improvements in voice communications over the standard helmet by simple improvement in the speech signal-to-noise ratio.

Recent technological advances have made application of the ANR practical. ANR is a means used to reduce noise levels in a personal hearing protector by measuring the noise in the earcup and reinserting a processed and out-of-phase noise signal back into the earcup through an earphone. The reinserted sound signal combines with the noise originally measured and causes it to be canceled. This out-of-phase canceling technique usually is very effective for low frequencies, below 800 Hertz, but generally is ineffective for higher frequencies. In some designs, the ANR device increases the noise level inside the earcup in the region where ANR crosses zero attenuation. Total protection provided by the ANR system consists of the passive hearing protection provided by the earcup, and the ANR noise reduction provided by the electronic system.

The CEP is a device which incorporates a miniature earphone coupled with a replaceable foam earplug tip, and may be used to improve hearing protection and speech commnications. [2] It can be worn in combination with the aviator's helmet providing protection similar to when using the yellow foam plug. The device consists of a miniature receiver encapsulated in a plastic housing, which includes a threaded adapter used for attaching the replaceable earplug. The earplug tip has an internally threaded insert channel that extends through the center from the base to tip, and mates with the threaded adapter on the transducer housing, shown schematically in Figure 3. The speech signal is delivered directly from the receiver into the occluded portion of the ear canal. The small wire used to connect the CEP into the communications system is highly flexible for comfort and small enough to reduce the potential for leakage when the wire is routed between the

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

earseal and the wearer's head. [3] This approach provides sound attenuation and speech intelligibility as good as any technique observed to date.

2 Discussion

Both techniques have been shown to reduce noise at the wearer's ear and improve the speech intelligibility characteristics of the aviator's helmet system. A study to determine the effect of these techniques on speech intelligibility for 20 normal and 20 hearing-impaired aviators was completed. Results of the study showed significant improvements over the standard helmet for both groups. Audiometric means of the two subject groups are shown in Figure 4. Speech intelligibility of the hearing-impaired aviators wearing CEP or ANR was compared with the 95 percent confidence interval for the normal aviator wearing the SPH-4 helmet, shown in Figure 5. The hearing-impaired aviators improved from 1 percent while wearing the SPH-4 to 65 percent while wearing the CEP helmet, and 40 percent while wearing the ANR helmet. The results of the study also showed that asymptotic levels of speech intelligibility are reached at much lower speech levels with ANR and CEP, as shown in Figure 6. The net effect should reduce speech levels required for communications and, therefore, reduce the hazardous effects of the speech signal. During field trials we found the intercommunications volume controls are reduced significantly from levels normally used for the standard helmet. [4]

considered when making a fielding decision. The areas concerning performance and safety are of primary importance. While user acceptance and cost may be of secondary importance, they are critical to the decision process. Safety must be considered, not only for the auditory performance enhancements, but for other mechanical factors designed to protect the aviator during normal missions and during unexpected or unplanned events. [5] Side impacts in the helicopter environment have been shown to produce significant head injuries during crashes and, in many cases, are preventable with energy-absorbing earcups. Figure 7 shows results of impact evaluations in the earcup of three ANR systems. The weight of the helmet is a significant factor for increased injury during a crash, and adds to the burden supported by the aviator during flight, as shown in Figure 8. The helmet has become a platform for many weapons system devices which are coupled to the aviator. This adds to the burden supported by the aviator, and techniques to reduce that burden must be explored.

Fielding considerations must include all aspects of how the user wears the helmet system and how various wearer configurations affect the performance of the system. For example, the ANR system is typically installed in a circumaural device, so the effects of equipment which compromise the earseal must be considered. CB protective hoods used by U.S. Army personnel are placed between the head and earseal and cause a significant loss in performance of the protective and communication characteristics of the helmet system. The effects of other ancillary equipment, such as spectacles, are important to the issue of the compromised earseal.

During the past year, USAARL has evaluated ANR systems manufactured by three U.S. corporations. The systems were provided to the Army under a cooperative research and development agreement for proposed laboratory and field testing. The ANR systems were compared to the standard helmet and to the CEP. Laboratory evaluations included the measurement of sound attenuation and speech intelligibility using 18 normal hearing flight students. The laboratory study included an evaluation of the effects of ancillary equipment, CB masks, and spectacles when used with the helmet. Field tests included questionnaire-based assessments completed by aviators after flying normal missions while wearing the test helmets. Assessments were accomplished in a variety of U.S. Army aircraft, to include the UH-60, OH-58, CH-47, and UH-1.

Results from the laboratory study conducted at USAARL show ANR and CEP produce improvements in speech intelligibility and sound attenuation when compared to the standard helmet. Figures 9 through 14 show results of sound attenuation measurements conducted on the test devices. Measurements for the insert devices, E-A-R and CEP, were conducted using ANSI S12.6, "Method for Measuring the Real-Ear Attenuation of Hearing Protectors," [6] while ANR devices were measured using MIL-STD-912, "Physical Ear Noise Attenuation Test." [7] Decreased sound attenuation or speech intelligibility performance when wearing spectacles with ANR or the standard helmet is minimal. However, wearing the CB mask causes significant reduction in the helmet system performance for the standard and ANR helmet systems. Small effects were os to be
protection provided by the CEP and the yellow foam earplug.

Speech intelligibility measurements were conducted using a wideband reproduction system to provide the speech material to the test device. Speech material consisted of single talker, commercially recorded W-22 word lists. Words were presented to the subject wearing the test device in a sound field of 105 dBA, simulating a UH-60 flying at 120 knot cruise. The test devices and word lists were counterbalanced to reduce learning effects. Results shown in Figures 15 through 17 compare performance of the test devices for each of the ancillary device combinations. Due to inadequate attenuation provided by the two ANR systems and the HGU-56/P helmet, the ambient noise in the test chamber was reduced 10 dB for these devices, while the yellow foam earplug and CEP were held at 105 dBA ambient noise.

While the speech intelligibility for the helmet when worn alone shows little effect, the loss of attenuation while wearing the mask is very significant. The loss of adequate communication with increased noise exposure, while compromising the visual system by wearing the CB mask, leaves the aviator in an uncertain state. Adding night vision goggles to the helmet system further complicates the situation.

Impulse noise hazard becomes an issue when considering the large number of rounds fired from open cockpit aircraft with weapon muzzles located near the crewmember's ear. ANR systems do not show any effect on reducing impulse noise levels encountered in the Army noise environments. Because of the high potential hazard to hearing, insert protection in combination with the helmet has been recommended for training scenarios involving weapons' fire from open cockpit aircraft.

The field evaluations were completed at three separate Active

Army units. The aircraft types used were the OH-58D, UH-1, UH-60, CH-47, and OH-6. More than 40 aviators participated, wearing each helmet system for a period of 1 week during normal mission scenarios. At the end of each week, they completed a questionnaire about the device they had worn. At the end of the study, they completed a questionnaire that covered all the test devices. The objective was to assess the users' helmet system preferences and solicit their judgment as to operational effectiveness.

At the beginning of the field test, one ANR system was removed from the test because it did not meet the safety requirements. The system did not provide communications capability during loss of battery power. The remaining two ANR systems, along with the standard helmet and the CEP, were included in the evaluation. Results of the evaluation, shown in Table 1, show the CEP and ANR systems provided subjective improvements over the standard helmet for noise reduction and speech clarity. Comfort was considered comparable for all of the helmet systems. Donning of the CEP was considered more difficult since it included an additional step in the process. Previous studies, along with this study, indicate about 80 percent of the U.S. Army aviators normally wear earplugs in combination with the helmet, which may account for the acceptance of the CEP system. The aviators did not feel any of the helmet systems reduced their awareness of the operational noises needed to ensure proper operation of the helicopter. In some cases, instability of the ANR circuitry was annoying but did not detract from successful mission completion. For overall preference, aviators favor the CEP over the other helmet systems.

3 Conclusions

ANR and CEP have reached the decision point in their development process and show promise for near term fielding. Besides the selection factors shown in Table 2, there are others which should be considered. Cost of aircraft modification, helmet system cost, logistics, and reliability should be evaluated carefully when considering the use of ANR or CEP in the helicopter environment. It is the authors' opinion that the CEP approach provides the best solution for all aspects of hearing protection, auditory performance, and many other areas of consideration.

4 References

1. Department of Defense Instruction 6055.12, 26 Mar 91. Hearing Conservation Program.

2. Ribera, J.E., Mozo, B.T., Mason, K.T., and Murphy, B.A. 1996. Communication and noise hazard survey of CH-47D crewmembers. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory. USAARL Report No. 96-02.

3. Mozo, B.T., Murphy, B.A., and Ribera, J.E. 1995 User acceptability and comfort of the Communications Earplug (CEP) when used in the UH-1 helicopter. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory. USAARL Report No. 95-17.

4. Ribera, J.E., Mozo, B.T., and Murphy, B.A. 1996. Subjective Evaluation of the Communications Earplug with Flexible Harness (CEP/FH) among CH-47D crewmembers. Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory. USAARL Report No. 96-29.

5. Shanahan, D. F.,Basilar skull fracture in U. S. Army Aircraft Accidents, USAARL Report No. 85-11, Ft. Rucker, AL, U.S. Army Aeromedical Research Laoboratory, 1985.

6. American National Standards Institute. 1984. Method for the measurement of the real-ear attenuation of hearing protectors. 12.6 1984.

7. Department of Defense. 11 December 1990. Physical-ear noise attenuation test. MIL-STD-912.

Acknowledgements: This work was funded in part by the U.S. Army Project Managers Office, Aircrew Integrated Systems. The authors are indebted to Ms. Barbara Murphy who did most of the data collection and data entry.

Table 1. Mean results of operational assessment. Rank ordered for 1='BEST' to 4='Worst'.

Test Device	Speech Clarity	Noise Reduction	Donning	Comfort	Outside sounds	Stability	Preference (Percent)
HGU-56/P	3.6	3.6	1.4	2.3	3.4	2.4	5
ANR1	1.9	1.9	2.4	2.1	2.6	2.3	33
ANR2	2.8	2.6	2.5	2.8	2.5	2.7	5
CEP	1.7	1.9	3.2	2.6	1.2	2.5	57

Table 2. Factors for consideration during the selection process.

FACTOR	ANR	CEP
Cost:	\$450.00-\$1750.00	<\$100.00
Added Weight:	(90 to 312 gm)	(-28 to 11 gm)
Aircraft modification Cost:	\$1000-\$5000	Not Required
Compatibility:	Reduced Performance	Unaffected

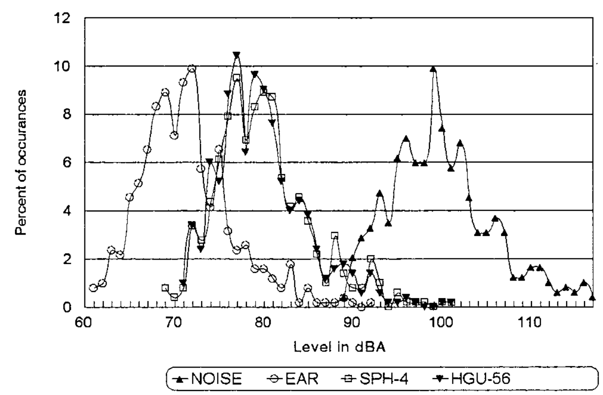


Figure 1. Distribution of noise exposure in U. S. Army Aviation

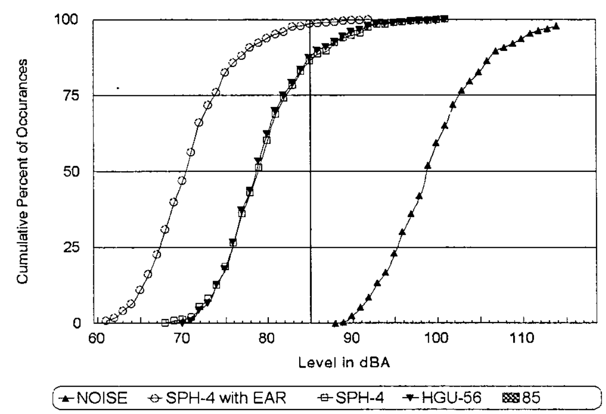


Figure 2. Distribution of noise exposure in U. S. Army Aviation

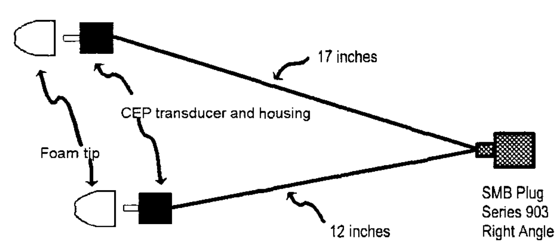


Figure 3. CEP unit schematic diagram

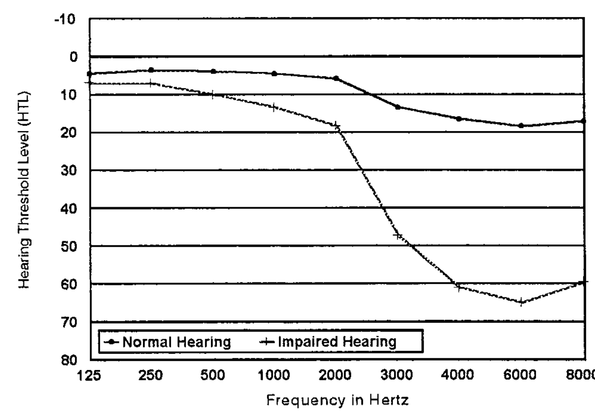


Figure 4. Mean audiometric thresholds of normal and hearing impaired aviators

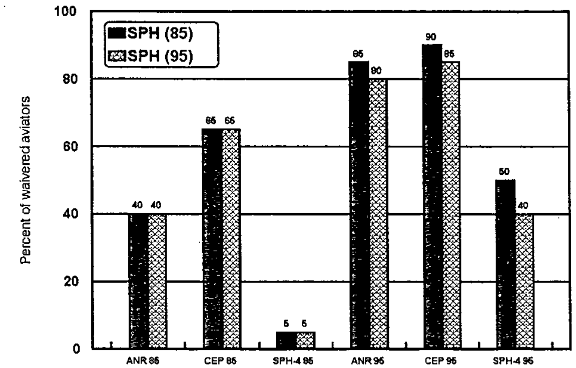


Figure 5. Speech intelligibility improvement of hearing impaired aviators compared with 95 percent confidence interval of normal aviators wearing SPH-4

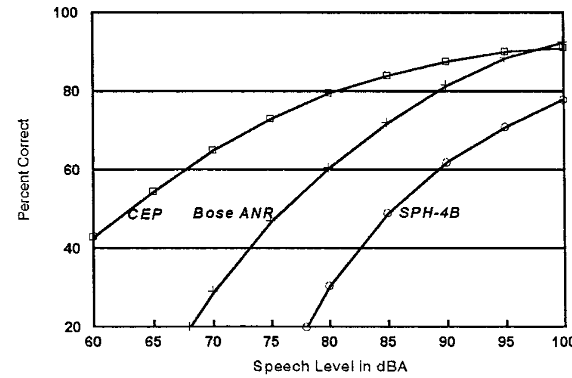


Figure 6. Speech intelligibility characteristics in noise simulating UH-60 at 120 knot cruise

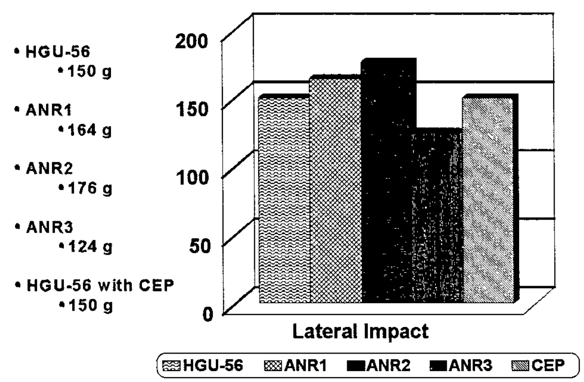


Figure 7. Lateral impact characteristics for CRDA test helmets

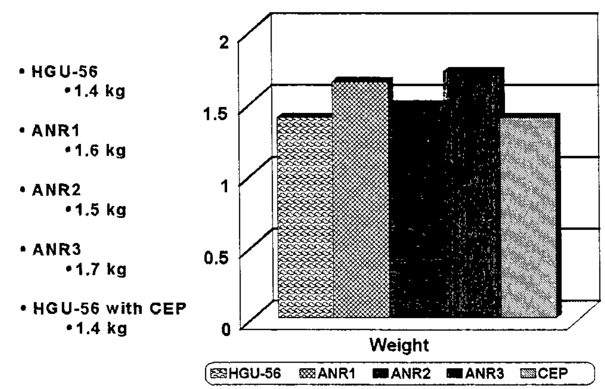


Figure 8. Weight of CRDA test helmets

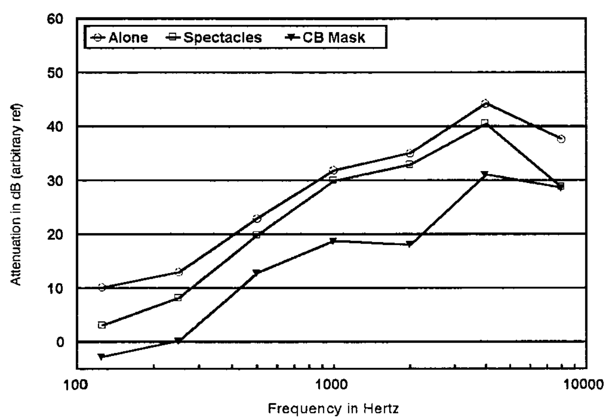


Figure 9. Sound attenuation of the HGU-56/P helmet

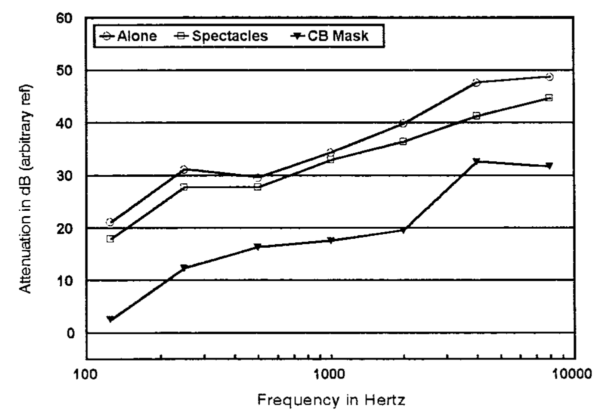


Figure 10. Sound attenuation of ANR1 helmet

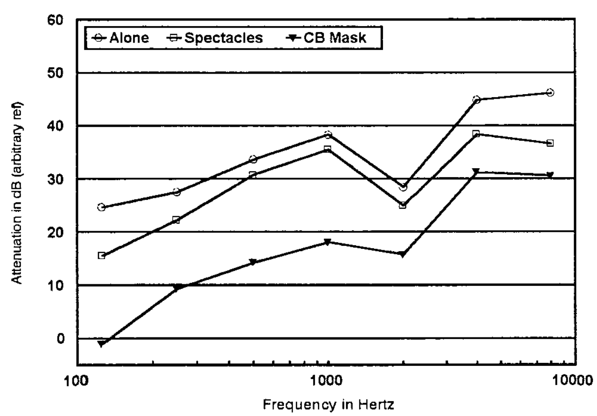


Figure 11. Sound attenuation of ANR2 helmet

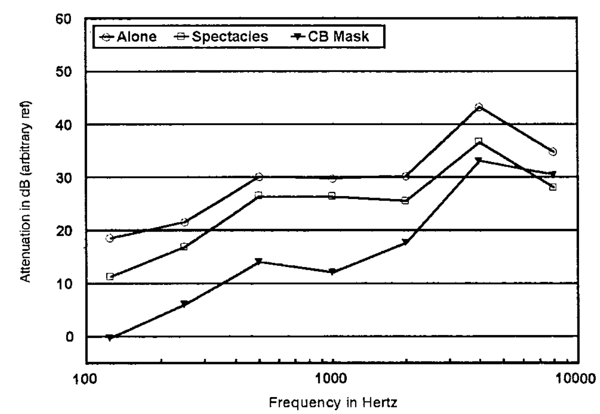


Figure 12. Sound attenuation of ANR3 helmet

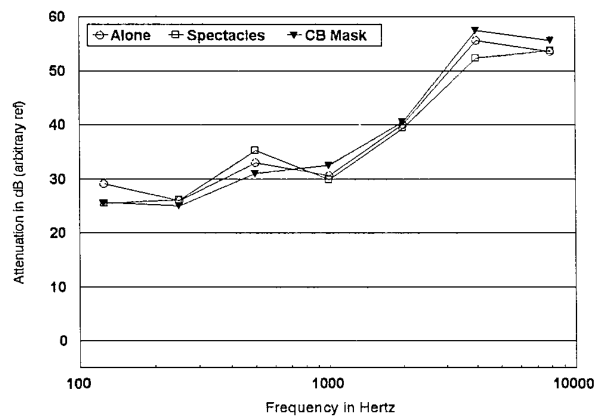


Figure 13. Sound attenuation of HGU-56/P with CEP

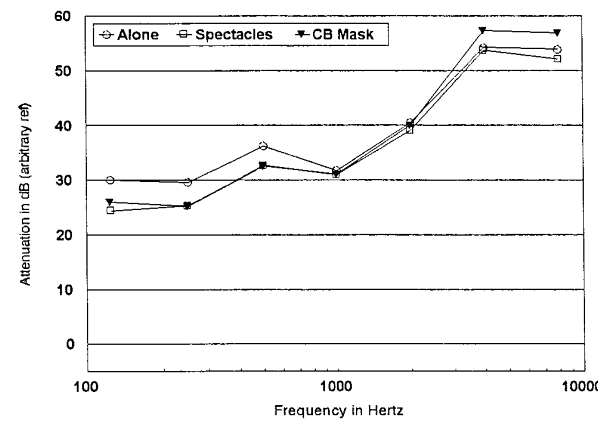


Figure 14. Sound attenuation of HGU-56/P with foam earplug

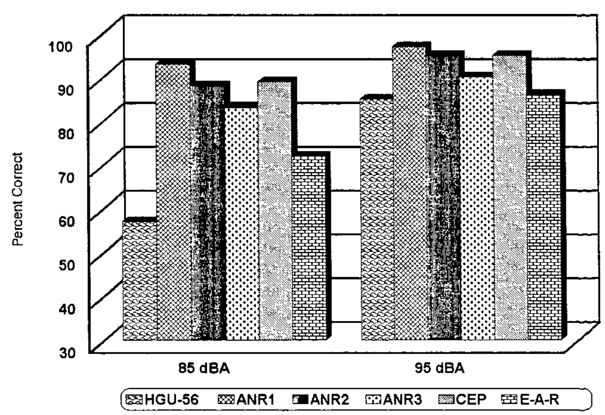


Figure 15. Speech intelligibility when worn alone

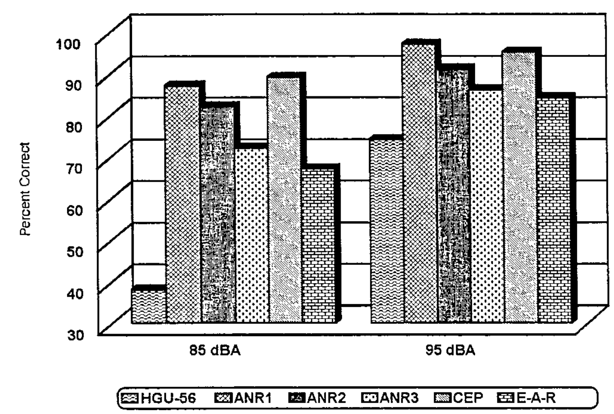
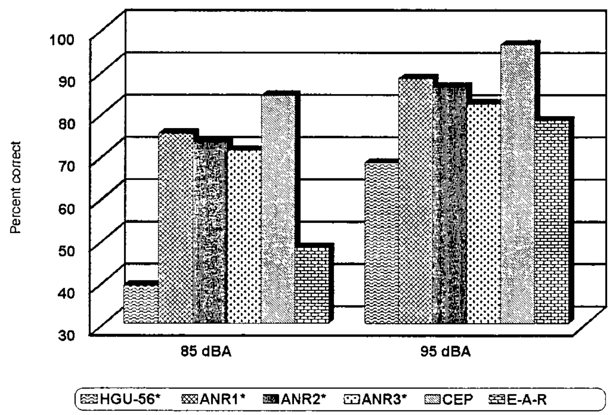


Figure 16. Speech intelligibility when worn with spectacles



*Ambient noise at 95 dBA.

Figure 17. Speech intelligibility when worn with CB Mask.

CONSTRAINTS IN THE APPLICATION OF PERSONAL ACTIVE NOISE REDUCTION SYSTEMS

R.B. Crabtree
Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West, P.O. Box 2000
North York, Ontario, Canada, M3M 3B9

1. SUMMARY

Active Noise Reduction (ANR) systems built into personally-worn headsets and helmets, when properly designed and carefully fitted, have shown considerable potential for reducing noise exposure and improving the listening conditions under which auditory tasks are carried out in military operations. Performance limitations have been identified in certain devices, however. Some have a tendency to overload easily or to cease operating under adverse conditions, and others become unstable when the seal around the ear is broken.

Recent findings indicate strongly that proper fitting around the ear is a functional necessity for satisfactory ANR operation. This is particularly true of units having a low tolerance to overloading and those which continue to operate in the infrasound frequency range. As a consequence, the function of any ANR system must be understood within the context of its intended operating environment in order to estimate whether the system will perform satisfactorily.

2. BACKGROUND

Personal Active Noise Reduction is an electro-acoustic technique for promoting the partial cancellation of sound within the ear cup of a hearing protector. Operating at frequencies below 1000 Hz, ANR is capable of providing greater attenuation at low frequencies than can be obtained by conventional (passive) means. The benefit is greatest in environments containing substantial amounts of low-frequency energy, such as helicopters and tracked vehicles.

DCIEM is studying ANR system characteristics to aid in selecting commercial devices best suited to applications in Canadian Forces environments. Early work consisted of evaluating attenuation properties by objective (physical-ear) and subjective (loudness balance and masked threshold methods) and studying signal detection capabilities among other attributes (Refs. 1 and 2). A recent outgrowth of this work has been the development and implementation of a number of additional tests with which to assess specific ANR characteristics. These include the saturation threshold (the limiting sound level in which systems continue to function properly), issues of fitting integrity, speech discrimination in active and passive modes, and general suitability (freedom from instability, heat build-up and fitting discomfort). The saturation threshold and the role of fitting form the principal subject matter of this paper.

Our laboratory work has confirmed that personal ANR can facilitate the successful execution of listening tasks at lower levels of presentation than would otherwise be necessary. One such listening task involves the detection of auditory signals, for example, sonar returns in maritime helicopter operations. An early study comparing active and passive systems showed that sine wave tone bursts could be detected by observers at significantly lower levels of presentation in a background of simulated helicopter noise when ANR was used (Ref. 1). What was not anticipated was that signal detection capability would be enhanced at frequencies well above the ANR operating bandwidth as shown in Figure 1, an outcome which could not have been predicted on the basis of active attenuation performance alone.

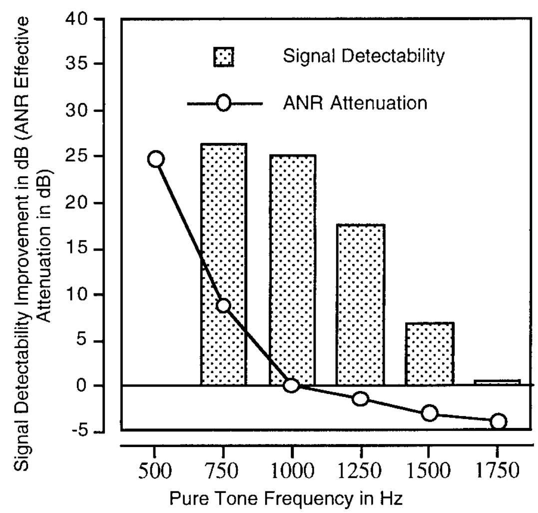


Fig.1. Improvement in Absolute Signal Detectability due to ANR Attenuation in a Background of Simulated Sea King Noise

It was presumed that the signal detection performance of the experimental observers was improved through the control of upward spread of auditory masking. This psycho-acoustic effect, sometimes called forward masking, refers to the capability of low-frequency high-amplitude sound energy to mask or hide sounds occurring at higher frequencies. Although most studies on forward masking have used mid-frequency tones or band-

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

limited noise as maskers (Refs. 3 and 4), there is also evidence that infrasound of sufficient amplitude is able to produce a similar effect (Ref. 5). An ongoing assessment of the noise levels at the ears of flight personnel using passive flight helmets in rotary-wing aircraft has shown that the intercom and radio sound levels are invariably higher than those considered necessary for adequate speech discrimination. This observation supports the probable occurrence of forward masking in real environments. It seems evident that the reduction of forward masking is one of the most important capabilities which can be attributed to ANR.

In flight environments, there is a tendency for ANR performance to be compromised through a combination of factors, including fitting difficulties with consequential sound leakage, and electro-mechanical limitations in the equipment itself which can lead to distortion, saturation or instability (Ref. 6). Anecdotal evidence suggests that these effects are commonly noted by flight crew, in spite of the potential of ANR to ameliorate forward masking through noise attenuation when properly applied. Consequently, there is a need to understand the low-frequency performance of personal ANR equipment and the factors on which performance depends, so that this emerging technology may be applied to the best possible advantage.

3. NEW STRATEGIES TO EVALUATE ANR

In recognition of these requirements, recent research at this laboratory has been aimed at developing a series of measurement strategies intended to isolate and quantify the factors contributing to the low-frequency performance of commercial ANR equipment (Ref. 7). One of the techniques to be described is a method which determines the limiting at-ear sound levels in which any given device will continue to function properly, and another assesses the attenuation properties of the ear shell and ear cushion in isolation from the effect of ANR. In the third, a technique is described in which overall attenuation performance is measured under both ideal and less-than-ideal conditions. In any selection process, these factors are considered in combination with the results of speech discrimination tests, an assessment of general suitability, and the nature of the sound field in which the system is to be used.

Measurement of Saturation Threshold

ANR system exposure to very high-amplitude low-frequency audible and sub-audible sound can lead to saturation or overload of the ANR electronics. This causes the system to generate extraneous noise at the ear, described variously as a clicking, popping or oil-canning sound. A technique was developed whereby the threshold or onset of overload could be determined experimentally through the direct excitation of the air volume confined within the ear cup of an ANR device.

A KEMAR acoustical mannequin headform with Zwislocki coupler artificial ears (Ref. 8) is modified by removing the couplers and mounting plates from the ear cavities. This provides 27 mm circular openings from the circumaural areas into the hollow headform. Calibrated 13 mm microphones are suspended in these openings such that their diaphragms are flush with the outer surface of the headform, allowing air to pass freely through the openings.

The hollow neck of the headform is attached to an opening in a loudspeaker enclosure containing a 200 mm low-frequency driver. Since there are no other openings in the enclosure, the back wave is acoustically coupled to the interior volume of the headform. When an ANR device is placed over the ear openings and the loudspeaker is driven by a very low-frequency pure tone, it is possible to excite the ear cavities to sound pressure levels exceeding 140 dB. The ANR system cannot distinguish between this type of excitation and that which would normally permeate the ear shells, thus the system under test will attempt to establish an opposing noise field. Since the measurement microphones are placed in proximity to the cancellation transducers, they "hear" the onset of distortion or overload in the form of extraneous noise. This is clearly audible over headphones used to monitor the microphones as the excitation sound level is varied in the region of the threshold. Alternatively, the onset of distortion can be monitored by connecting a signal analyzer to the measurement microphones and observing the rising pattern of sound energy above the excitation frequency as the threshold is exceeded.

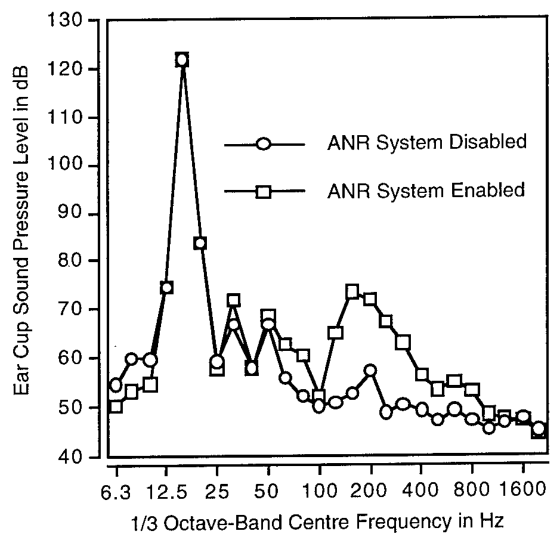


Fig.2. Typical Extraneous ANR System Noise Resulting from Presenting a 16-Hz Pure Tone at a Level 5 dB above the Saturation Threshold

A typical noise spectrum resulting from overload is shown in Figure 2, where the excitation is a 16-Hz pure tone presented

5 dB above the saturation threshold. For reference, the comparable spectrum with ANR defeated is also shown. The difference between the traces represents the extraneous noise which in this case is most pronounced at frequencies between 100 and 1000 Hz. This effect raises the possibility of interference with the lower portions of the speech band, and helps to explain the negative observations reported by flight personnel.

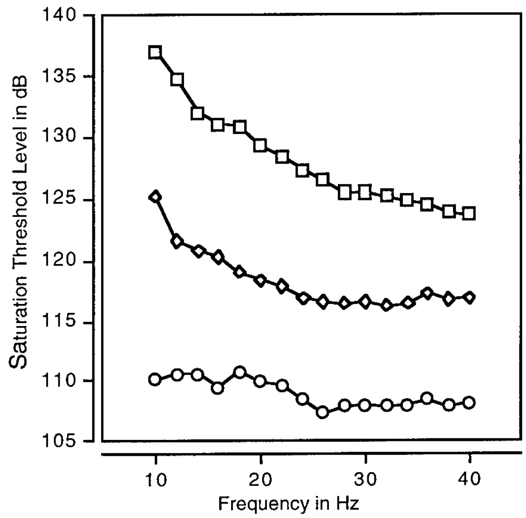


Fig.3. Saturation Thresholds for Three Typical ANR Systems

Whichever method of detection is used, the sound pressure levels registered by the microphones at the saturation threshold can be plotted as a function of frequency, as shown in Figure 3. A tendency towards lower thresholds is thought to be attributable to two compounding factors. Firstly, those systems capable of providing significant cancellation within this frequency range simply work hard in the presence of infrasound excitation. Secondly, hardware constraints such as the excursion limits of the cancellation transducers or the available drive power restrict the size of cancellation signal that can be generated.

Although a high overload threshold may indicate that the unit is particularly robust, it may also show that it is simply insensitive to energy in this frequency range. A review of the unit's active attenuation performance within the same frequency range will aid in making this distinction. It is noteworthy, however, that the devices having high cancellation capability in the infrasound region are perceived by the user as creating the best listening environment when operated below the overload threshold, in spite of the relative insensitivity of the human hearing system at these frequencies.

Ear Shell and Ear Cushion Attenuation

In any ANR system, the active attenuation due to electronic assist is complemented by the passive reduction provided by the ear shell assembly. Although the measurement of saturation threshold describes ANR behaviour within the ear cup at very low frequencies, it does not quantify the effect of the ear cushion or the seal against the side of the head. This information is needed to estimate the magnitude of an external sound field which will cause the overload threshold to be exceeded.

To study passive reduction capability, the entire ANR system is subjected to a low-frequency high-pressure sound field. The pressure vessel in which this test is carried out comprises a large sealed loudspeaker enclosure with a 300 mm driver. The driver is removable for access to the interior of the enclosure, which contains a heavy flat-plate coupler having a measurement microphone at its centre. The coupler is used to carry out insertion loss measurements of attenuation. A noise spectrum is obtained from the microphone as the coupler is pressed against one of the ear cushions and another is obtained with the cushion and coupler separated from each other, as the driver is excited by low-frequency pink noise. Passive attenuation is taken as the difference between the two spectra.

Experience has demonstrated (Ref. 9) that the fit of a protective device rarely approaches the quality of that obtainable against a flat plate coupler, thus it is acknowledged that the attenuation data obtained in this way should be thought of as ideal. It was therefore considered prudent to study the effect of a less-than-perfect seal to the coupler.

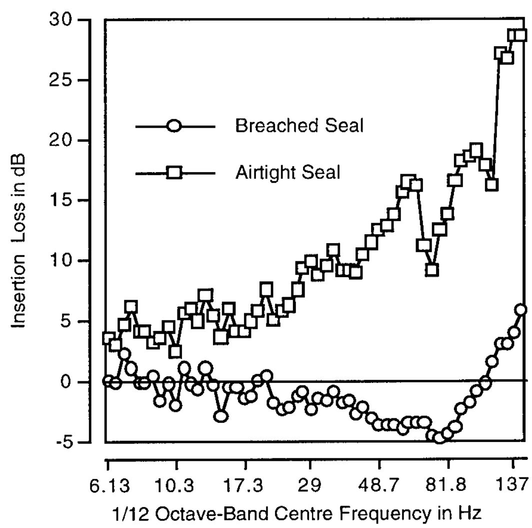


Fig.4. Sample Ear Cup / Ear Seal Insertion Loss with ANR Electronics Disabled

For this purpose, a tube 1.6 mm inside diameter x 25 mm length is inserted between the ear cushion and the coupler surface. It is embedded in a wedge of plasticine to prevent air movement around its periphery. The size of the tube was chosen to approximate the leakage cause by the insertion of a metal side frame of Canadian Forces issue sun glasses under the seal. In the absence of additional leakage, this represents a relatively small acoustical path.

The results of a typical low-frequency insertion-loss measurement with ANR in passive mode are shown in Figure 4. The upper trace shows the attenuation achieved with an ideal (airtight) seal against the flat-plate coupler containing the measurement microphone, and the lower trace the effect of breaching the seal into the ear cavity by means of the tube described above. The leakage path appears to act as a resonator with the enclosed air volume which amplifies sound energy in the 50 - 100 Hz region and generally nullifies any attenuation below 30 Hz.

Thus, for a given ANR system to perform satisfactorily in this helicopter, it needs to be capable of generating very high levels of infrasound.

Overall Performance Characteristics

The two techniques described above allow the separation of ANR function from the determination of low-frequency passive attenuation provided by the ear shell structure. One can also assess overall system performance by testing in an environment closely duplicating the noise in which the equipment might be used, for example to estimate hearing hazard. DCIEM has developed a noise simulation facility fully meeting these requirements, in which interior noise recorded in aircraft and in ground vehicles may be faithfully reproduced, in level, in bandwidth and in temporal pattern. The result is a test environment closely duplicating the actual noise encountered in the field. This capability allows testing under controlled and repeatable conditions and substantially lessens dependence on field trial resources for routine testing.

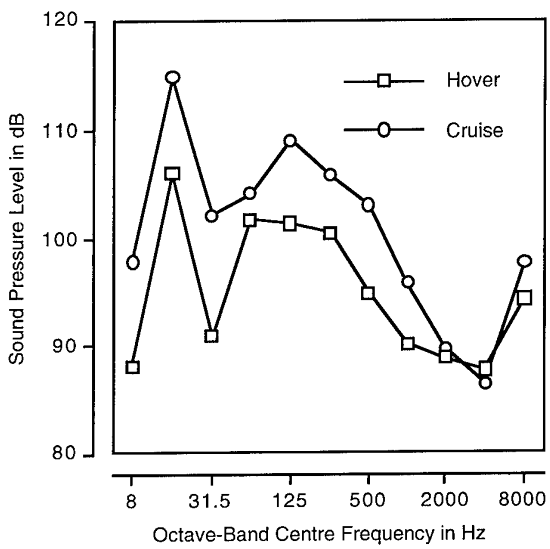


Fig.5. Ambient Sound Levels on the Flight Deck of the Sea King Helicopter

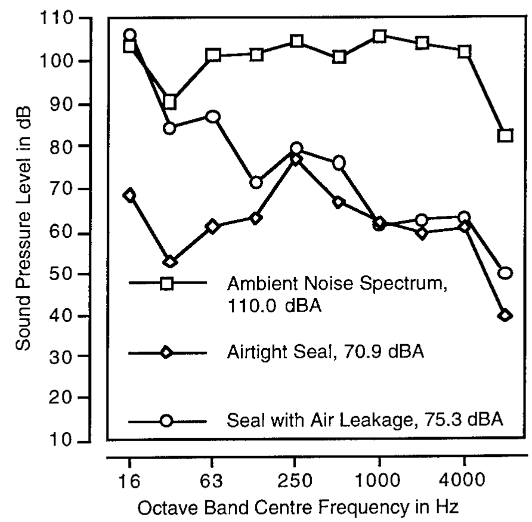


Fig.6. Attenuation Characteristics of Effective ANR System when operated against Flat Plate Coupler in Realistic Sea King Helicopter Noise

Acoustical leakage of this magnitude can be particularly troublesome in operational environments containing low-frequency sound energy. For example in the Canadian Forces Sea King maritime helicopter, the largest acoustical input to the cabin occurs at the main rotor blade-pass frequency, about 17 Hz, as shown in Figure 5. An air leak as small as that described above would force the ANR electronics to accommodate ambient (rather than attenuated) levels of infrasound, as well as higher-than-ambient levels of the 2nd - 5th order harmonics of rotor blade pass noise.

The results of some physical measurements carried out in this facility are given in Figures 6 and 7. The overall flat plate sound levels inside the ear cup of an effective wide-range ANR system is shown in Figure 6, together with the Sea King helicopter excitation spectrum measured via the coupler while separated from the ANR system. The difference between these curves represents the total attenuation of the device, that is, the passive attenuation as augmented by the action of ANR. Shown also is the dramatic effect of breaching the seal against the flat plate by means of the small tube described earlier. Notably, the

performance decrement is considerably larger than would be predicted solely by the loss of passive attenuation.
Clearly, leakage of this magnitude seriously compromises the intended effect of ANR.

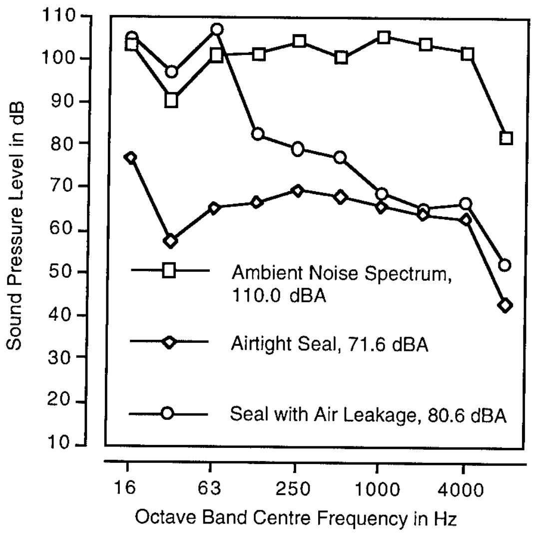


Fig.7. Attenuation Characteristics of Less Effective ANR System when operated against Flat Plate Coupler in Realistic Sea King Helicopter Noise

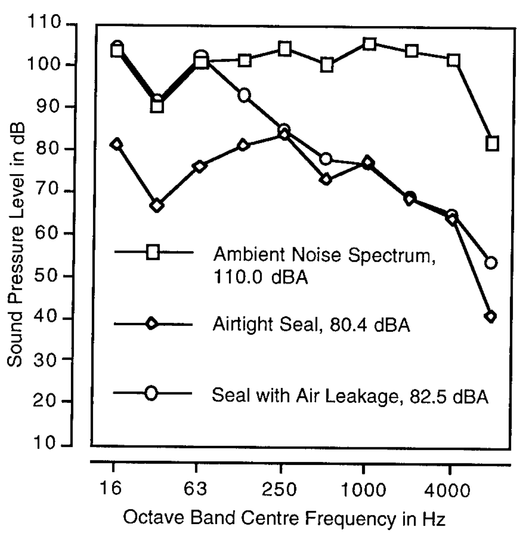


Fig.8. Attenuation Characteristics of Passive Communications Headset when operated against Flat Plate Coupler in Realistic Sea King Helicopter Noise

In Figure 7, the performance of a less effective system is shown for a measurement carried out under identical conditions. In both instances, there is general tendency for the leaky condition spectrum and the excitation spectrum to become asymptotic at low frequencies. This also occurs in devices not equipped with ANR, for example in Figure 8, although as expected, this headset provides less protection under airtight conditions than either of the active protectors.

4. IMPLICATIONS

Electro-Acoustic Limitations

It has been shown that commercial implementations of ANR differ considerably in their ability to operate effectively at very low frequencies. Built-in filter characteristics in some units permit operation in the infrasound region, usually with relatively low tolerance to overload, while others are relatively insensitive to infrasound. Subjectively however, the units which "sound best" are those with extended operating bandwidths, particularly when used in helicopter environments. Manufacturers are constrained by the size, excursion capability and power dissipation of transducers built into the ear cup. Larger, more powerful units would lessen the tendency to overload, particularly in the presence of sound leakage, but would remain restricted in providing additional attenuation because of their fixed filter and gain characteristics. A system capable of adaptation such as the one being developed by the National Research Council (Ref. 10) should lessen the overall dependence on effective fitting.

Fitting Limitations

The difficulties associated with adequate fitting of hearing-protective devices in field environments has been a significant health concern for many years (Ref. 9). Our own studies have indicated that the reception of radio and intercom messages is a significant component in the acquisition of noise dose, and improper fitting of helmets or headsets invariably requires higher intercom levels for adequate speech discrimination. Fitting difficulties are no different with ANR devices, except that the consequences are more severe in terms of loss of attenuation (see Figures 6, 7 and 8). Results such as these underscore the crucial importance of fitting, yet clearly indicate the performance potential achievable under ideal conditions. More work is required to assess performance on real subjects such that the inevitable decrement in performance due to fitting anomalies may be better understood.

5. SUMMARY

Environments in which ANR has the potential to provide the greatest benefits to the user often contain low-frequency noise of sufficient amplitude to cause ANR equipment to malfunction. ANR performance at very low frequencies appears to depend upon the capability to generate cancellation waveforms within this frequency range, upon hardware constraints such as transducer excursion limits and upon the integrity of the seal against the head. The data presented in this paper emphasize the

importance of understanding the behaviour of ANR devices at extremely low frequencies and the relationship to the noise environment in which they will be used.

6. ACKNOWLEDGEMENTS

The author is grateful to Mr. Andrew Welker of Welker Audio Consulting for his assistance in developing and applying the pressure simulators described in this paper, and for collecting and analyzing the data. He also thanks Patricia Odell and Stephen King of DCIEM for their work in obtaining and preparing the field data.

7. REFERENCES

1. Forshaw S.E., Rylands J.M. and Crabtree R.B. Evaluating the Effectiveness of Active Noise Reduction in Flight Helmets. DCIEM 88-R-34, Department of National Defence, Canada, 1988.

2. Rylands J.M., Forshaw S.E. and Crabtree R.B. The Evaluation of Active Noise Reduction ANR in Flight Helmets and Headsets ASCC Advisory Publication 61/103/15, 1991.

3. Biljer R.C. and Hirsch I.J. Masking of Tones by Bands of Noise. J. Acoust. Soc. Am., Vol. 28, 1956.

4. Carter N.L. and Kryter K.D. Masking of Pure Tones and Speech. Journal of Auditory Research, Vol. 2, 1962.

5. Broner N. The Effects of Low Frequency Noise on People - a Review. Journal of Sound and Vibration, 1978, 58(4).

6. Crabtree R.B. Active Noise Reduction Systems; Their Interaction with Very Low Frequency Energy. Proceedings of the 12th Triennial Congress of the International Ergonomics Association, Toronto, Canada, August 1994.

7. Welker A.C. Evaluating Active Noise Reduction System Performance within Low Frequency Noise Environments. University of Waterloo Engineering Co-Op Student Work Report, DCIEM, Toronto, Canada, 1993.

8. Burkhard M.D. and Sachs R.M. Anthropometric Manikin for Acoustic Research. J. Acoust. Soc. Am. Vol. 58, 1975.

9. Berger E.H., Royster J.D., Casali, J.G., Merry C.J., Mozo B.T. and Royster L.H. Results of a Pilot Interlaboratory Comparison of REAT Measurements of Hearing Protectors. J. Acoust. Soc. AM. Suppl. 1, Vol. 88, 1990.

10. Pan G.J., Brammer A.J. and Crabtree R.B. Adaptive Active Noise Reduction Headset for Helicopter Use. Proceedings of AGARD Symposium on Audio Effectiveness in Aviation, Copenhagen, Denmark, October 1996.

Evaluation de casques à réduction active de bruit : protection auditive et intelligibilité

Assessment of active noise reduction hearing protectors : noise attenuation and speech intelligibility

L. PELLIEUX, D. SARAFIAN - *IMASSA-CERMA, BP 73, 91223 Brétigny sur Orge, France*
G. REYNAUD - *SEXTANT Avionique, BP 91, 33166 Saint-Médard en Jalles, France*

Résumé

La faible isolation acoustique des casques des pilotes d'aéronefs militaires entraîne à long terme un risque pour l'audition du personnel. Par ailleurs, la mauvaise intelligibilité des communications rend nécessaire une écoute à niveau sonore élevé, ce qui aggrave encore le risque.

Dans ce cadre, une évaluation de huit systèmes à réduction active de bruit, casques disponibles commercialement et prototypes en version casque intégral, a été effectuée.

L'originalité de la démarche réside dans la définition et la mise en oeuvre de protocoles expérimentaux prenant en compte l'étude du comportement des systèmes en utilisation non nominale, la mesure objective des atténuations passive et active par la méthode MIRE sur cinq sujets, la prédiction objective d'intelligibilité par l'attribution de l'indice STI, et son évaluation subjective par test CVC.

Summary

Hearing protection offered by current pilot helmets is far to be fully satisfying as shown by the large number of hearing losses observed in military aviators at retirement age. Due to the poor intelligibility of communication channels the sound volume has to be significantly increased which adds a dangerous auditory stressor.

Eight hearing protectors such as commercially available active noise reduction (ANR) headsets and prototype helmets, equipped with ANR earshells, were assessed in order to estimate their efficacy for both noise attenuation and improvement on speech intelligibility.

The assessment was based on original experimental protocols including abnormal conditions, objective measurement of both passive and active attenuations by the MIRE method, subjective prediction of intellligibility by measuring the Speech Transmission Index, and its subjective evaluation through CVC tests. Realistic jet and helicopter noisy environnments and a pink noise have been used to perform the tests. The results obtained with the various systems assessed are presented and discussed.

INTRODUCTION

La protection auditive, et de façon générale les équipements audio embarqués, constituent un secteur sous développé de l'avionique. Nombre de pilotes et de navigants soufrent à terme d'atteintes irréversibles. En effet, le niveau du bruit en cabine, très souvent supérieur à 100 dBA, est trop faiblement atténué par le casque et les coques des écouteurs, ce qui oblige les utilisateurs à augmenter le volume d'écoute de leur retour audio afin de rendre intelligibles les communications. A terme, cette démarche est bien entendu désastreuse pour l'audition...

Une nouvelle technique de protection au bruit est apparue, issue du principe de la réduction active de bruit (Active Noise Reduction, ANR). Ce procédé fait l'objet d'un brevet déposé en 1936 par le Docteur LUEG et a été décrit de façon approfondie en 1957 par OLSON et MAY. Les premières réalisations pratiques ne sont apparues que dans la dernière décennie : cela est principalement dû aux améliorations récentes des performances des transducteurs électroacoustiques (microphones à électret et haut-parleurs). La réduction active de bruit permet d'augmenter l'efficacité des protecteurs auditifs pour les basses fréquences (en deçà de 1 kHz), zone dans laquelle la protection passive est insuffisante, du fait de la faible masse surfacique des protections. Le principe (fig1) est fondé sur la superposition d'un signal sonore émis par une source secondaire et du signal primaire -le « bruit »- à supprimer : la somme de deux signaux en opposition de phase et de formes d'ondes identiques est nulle.

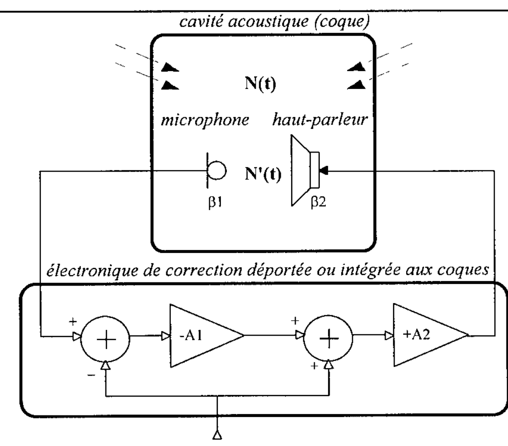


fig 1 *Principe de la réduction active de bruit*

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

L'objectif de l'étude ici rapportée est d'examiner les intérêts potentiels, en termes d'efficacité de la protection auditive et d'augmentation de l'intelligibilité, apportés par l'utilisation de casques combinant les protections passive (absorbants acoustiques) et active (dispositif électronique). Pour ce faire, des casques commercialement disponibles ont été sélectionnés, et des prototypes de casques « pilote » intégrant des écouteurs à réduction active de bruit ont été spécifiés et développés.

L'originalité de cette étude réside dans la définition et la mise en oeuvre de protocoles expérimentaux prenant en compte l'étude du comportement des systèmes en utilisation non nominale (bougé, retrait, coupure d'alimentation, forts niveaux sonores), la mesure objective des atténuations passive et active par la méthode MIRE, la prédiction objective d'intelligibilité par l'attribution d'indice STI, la restitution d'ambiance sonore « aéronef » réaliste.

Dans une première partie, les différentes méthodes d'évaluation des systèmes à réduction active de bruit sont rappelées, puis dans un second temps les moyens associés à la mise en oeuvre des méthodes retenues sont décrits. Les résultats des essais sont enfin présentés et discutés.

1 METHODES

1.1 ESSAIS EN SITUATION LIMITE

L'objectif initial de cette phase est d'identifier, par divers tests, les casques actifs dont le port peut se révéler dangereux pour les expérimentateurs. En effet, l'utilisation d'une protection auditive mettant en oeuvre un système actif électroacoustique bouclé impose une sécurité importante du dispositif : un effet LARSEN entre le haut-parleur et le microphone de capture serait catastrophique si près du tympan. Ce phénomène est malheureusement fréquent si les marges de stabilité du système sont trop faibles.

1.1.1 Comportement des casques en présence de fuites acoustiques :

Le but de cette phase est d'étudier la stabilité du système en présence de fuites acoustiques de diverses origines, i.e. s'assurer que le casque ne produit pas de régénération de bruit potentiellement dangereuse lors de « l'ouverture » de la cavité acoustique, constituée par l'oreille et l'écouteur. Les situations réalistes suivantes ont été reproduites, avec réduction active en service : bougé puis retrait complet des coques, chocs contre les coques. Ces cas se présentent lorsque l'utilisateur repositionne son casque, le retire ou bien lorsque ses mouvements de tête provoquent des chocs contre les parois environnantes, alors que le dispositif actif fonctionne.

1.1.2 Comportement des casques en présence d'une coupure d'alimentation électrique :

Dans cette situation, nous étudions le comportement des casques lors de la déconnexion accidentelle (simulée) de l'alimentation électrique du dispositif de réduction active de bruit. Il s'agit de vérifier que ce type d'événements ne génère pas de bruits de commutation audibles trop importants, voire des instabilités. Les bruits de commutation sont généralement brefs mais très désagréables, ce qui est intolérable dans le cadre d'une utilisation courante.

1.1.3 Comportement des casques à forts niveaux sonores (>120dB) sur tête artificielle

Cette simulation correspond à une situation réaliste pouvant être rencontrée dans des phases de vol particulièrement bruyantes telles que le décollage avec postcombustion, les évolutions sous facteur de charge, le vol à basse altitude. En effet, l'intégration d'un système actif implique son utilisation permanente quelle que soit la phase de vol. Par conséquent, la fiabilité de son fonctionnement doit être garantie à tout moment. Pour cela deux expérimentations sont menées, l'une au moyen d'un bruit rose pour des mesure d'atténuation, l'autre à l'aide d'un sinus basse fréquence pour des mesures de distorsion.

bruit rose 120dBA

L'objectif est d'analyser le comportement spectral global du dispositif (« pompage » éventuel, mise hors service, efficacité du filtre de correction). Le protocole expérimental est le suivant :
- mesure du champ acoustique au centre de l'enceinte, avec un microphone de mesure, en l'absence de la tête
- mesure du champ acoustique sous le casque, réduction active de bruit hors service
- mesure du champ acoustique sous le casque, réduction active en service
- vérification de l'homogénéité du niveau du champ acoustique dans l'enceinte, en présence de la tête, à l'aide d'un sonomètre.

sinus basse fréquence fort niveau

L'objectif est d'étudier le niveau de « saturation » du système, en d'autres termes son aptitude à générer un « contre-bruit » basse fréquence de façon efficace, sans distorsion nuisible à l'intelligibilité (i.e. inférieure à 20 %) ou à la sécurité (instabilité totale du système ; régénération de bruit dangereuse). Le protocole expérimental est le suivant :
- mesure du niveau et de la distorsion au centre du champ acoustique, à l'aide d'un microphone de mesure, en l'absence de la tête (validation de la qualité de restitution sonore du dispositif)
- mesure de la distorsion sur les deux premières harmoniques sous le casque, réduction active de bruit hors service, en fonction du niveau sonore et de la distorsion mesurée dans le caisson sur microphone de référence (mise en évidence d'éventuels phénomènes de résonance dus à la structure de la tête)
- mesure de la distorsion sur les deux premières harmoniques sous le casque, réduction active de bruit en service, en fonction du niveau sonore et de la distorsion mesurée sur microphone de mesure.

1.2 ATTENUATION

1.2.1 Généralités sur les méthodes classiquement utilisées

La mesure de l'atténuation apportée par un protecteur auditif constitue la principale (et souvent l'unique) caractérisation de ses performances. La méthode de mesure classiquement utilisée repose sur l'évaluation du seuil audiométrique à différentes fréquences (octaves de 125 à 8000 Hz) de plusieurs sujets, avant, puis après, la mise en place du protecteur auditif (cf norme ISO 4869).

Cette méthode se limite à la caractérisation des dispositifs antibruit passifs, car seule l'atténuation passive est linéaire par rapport à l'intensité du signal sonore à atténuer (excepté dans le cas de mesures lors de tirs d'armes -pouvant atteindre 190 dBSPL- qui peuvent provoquer des non-linéarités dues au déplacement des masses par l'onde de choc). En outre, l'aspect totalement subjectif de cette méthode ne garantit pas une grande fiabilité ; la méthode par test d'audiométrie est gourmande en temps et nécessite un grand nombre de sujets. L'utilisation d'un système électronique de réduction active de bruit condamne ce type de méthode car d'une part l'électronique associée génère un bruit de fond d'un niveau supérieur au seuil d'audition, et d'autre part l'atténuation active dépend directement du type de bruit (essentiellement du niveau dans chaque bande de fréquence) : en effet, de façon évidente, les performances et le comportement du filtre et des transducteurs mis en oeuvre varient en fonction de la fréquence et de l'amplitude du signal à traiter.

En raison des remarques précédentes, il est nécessaire d'évaluer les performances d'un système actif en ambiance bruyante à divers niveaux (par exemple, un bruit rose qui offre l'avantage d'être aisément reproductible) ou, mieux, en ambiance réaliste en restituant l'environnement sonore dans lequel le système sera utilisé de façon opérationnelle. Cette démarche nécessite l'enregistrement préalable du champ acoustique en milieu réel (cockpit) et sa restitution sous forme d'un champs diffus généralement obtenu à l'aide d'une chambre réverbérante.

Ces données environnementales étant posées, nous pouvons décrire trois méthodes, objectives et subjective, d'évaluation des dispositifs antibruit.

1.2.2 Nouvelles méthodes de mesure liée à l'emploi de la réduction active de bruit

Il est possible de distinguer les méthodes subjectives des méthodes objectives.

- méthode subjective par égalisation de sonie

L'expérimentateur muni du protecteur auditif actif est placé dans un champ acoustique diffus, dont l'intensité varie périodiquement (typiquement toutes les secondes) entre deux niveaux. Pendant la phase « niveau fort » (resp. « niveau faible »), la réduction active de bruit est en service (resp. hors service). Le sujet ne doit entendre qu'une faible différence entre les niveaux puisque le système actif n'atténue que le bruit du niveau le plus élevé. Durant la phase « réduction active hors service », l'expérimentateur a la possibilité de régler le niveau le plus faible de telle sorte qu'il ne perçoive plus de différence entre les deux états. L'écart entre les niveaux de bruit à l'extérieur de la cavité est alors égal à l'atténuation apportée par le dispositif actif. Le bruit utilisé est à bande étroite limitée au tiers d'octave. Les fréquences centrales sont présentées dans un ordre aléatoire pour éviter un effet systématique d'apprentissage. L'atténuation est donc donnée en fonction de la fréquence par tiers d'octave, ou encore, après calcul, par octave. Cette méthode sophistiquée requiert une automatisation de toute la chaîne de mesure et d'émission du bruit. En outre, certains systèmes génèrent des bruits de commutation lors du passage passif-actif, il est alors nécessaire de modifier l'électronique d'alimentation pour les supprimer, ce qui n'est pas toujours possible compte tenu de la structure des coques (électronique intégrée).

- mesure sur tête artificielle

Un tel outil de mesure présente de nombreux avantages de commodités, de fiabilité, de reproductibilité, mais par contre ne tient pas compte des disparités morphologiques humaines. Quelques essais ont été réalisés sur tête artificielle. Nous avons noté des problèmes d'amplification de bruit dans certains tiers d'octave ainsi que des problèmes d'étanchéité. C'est pourquoi à défaut de trouver une tête artificielle adaptée, nous avons opté pour une expérimentation sur sujets humains.

- méthode de mesure de bruit sous écouteurs utilisant le microphone miniature intégré (microphone de « boucle »)

Le casque étant placé sur la tête d'un expérimentateur, ce microphone capture le bruit résultant dans la cavité acoustique ; un analyseur de spectre traite le signal en temps réel, réduction active de bruit hors puis en service. Par soustraction des spectres généralement obtenus avec une résolution au tiers d'octave, nous calculons l'atténuation active apportée par le dispositif. Les résultats fournis par ce type de mesure surestiment les performances du dispositif actif car d'une part le champ acoustique sous la coque du casque n'est pas homogène aux basses fréquences (en raison des très faibles dimensions), et d'autre part le système est par nature conçu pour minimiser le niveau sonore à l'endroit où se trouve ce microphone,

- méthode MIRE : utilisation d'un microphone miniature additionnel placé à l'entrée du conduit auditif

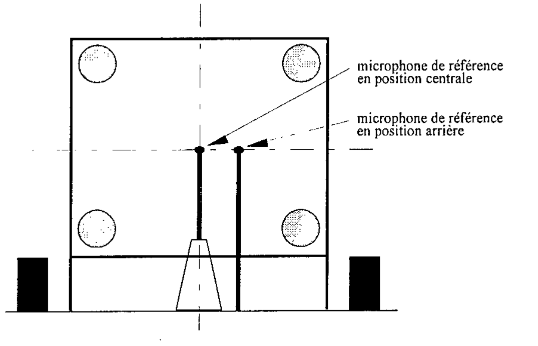
Cette méthode, communément appelée MIRE (Microphone In Real Ear), est devenue un nouveau standard international (ANSI S 312.42 1995 - ASA 116). Nous avons retenu et adapté cette méthode pour nos évaluations. Des études menées au TNO ont montré que les valeurs obtenues respectivement par la méthode MIRE et la méthode subjective sont très cohérentes (STEENEKEN 94).

1.2.3 Protocoles de mesure d'atténuation

Les opérations nécessaires à la mesure des atténuations sont les suivantes :

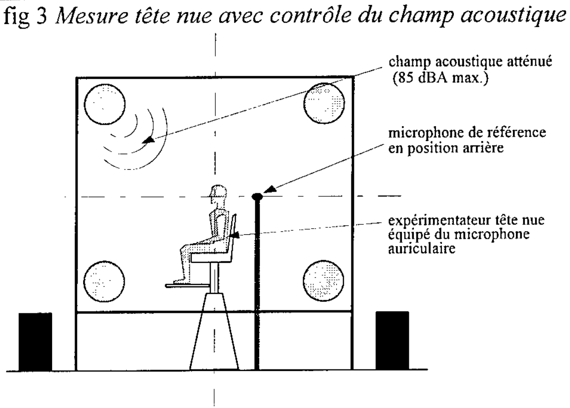
- 1) mesure du spectre de puissance caractérisant le champ acoustique restitué : à l'aide d'une perche, un microphone de mesure est placé au centre du dispositif de restitution sonore (fig 2).

fig2 *Mesure des champs acoustiques de référence*



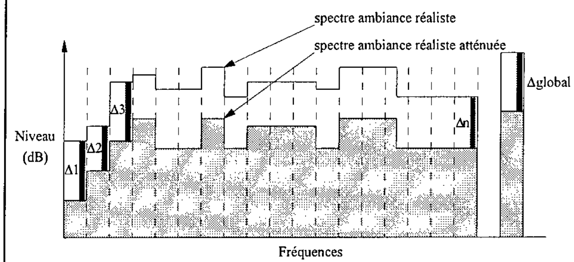
- 2) mesure du spectre de puissance du bruit réellement perçu par l'expérimentateur en l'absence de protection auditive. Bien évidemment, il n'est pas question d'exposer tête nue (fig 3) les expérimentateurs aux niveaux sonores des ambiances réalistes (>100 dBA). La mesure tête nue s'effectue donc à niveau atténué (80 à 85dBA) pour chacune des ambiances. Une évaluation préliminaire permet de connaître précisément les écarts dans chaque tiers d'octave entre niveau atténué et niveau réel (fig 4).

fig 3 *Mesure tête nue avec contrôle du champ acoustique*



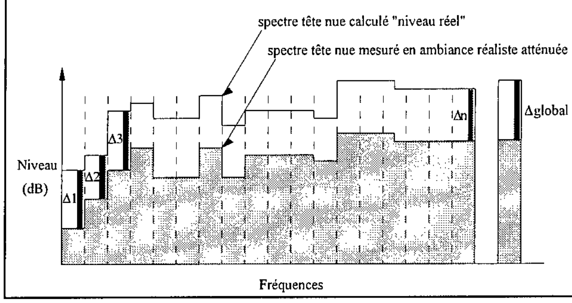
Pour obtenir le niveau que percevrait le sujet, il suffit de rajouter ces écarts à la mesure tête nue « niveau atténué » (fig 5). Grâce à cette méthode, nous prenons en compte les phénomènes d'amplification dus au pavillon et au conduit auditif qui se manifestent notamment aux fréquences élevées (cet effet est variable selon les individus), ce qui permet par la suite de mesurer des atténuations passives les plus réalistes possible.

fig4 *Mesure des imperfections de la chaîne de restitution sonore entre les niveaux réel et atténué*



N.B. : théoriquement, tous les écarts sont égaux.
Ce n'est pas parfaitement vérifié en pratique en raison des imperfections des hauts-parleurs.

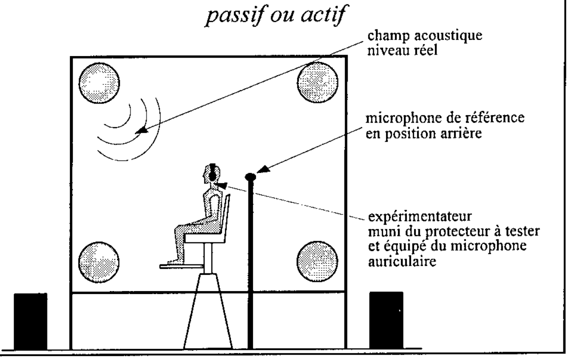
fig 5 *Reconstruction du spectre perçu par l'expérimentateur à niveau réel*



- 3) mesure du spectre de puissance du bruit sous le casque, réduction active de bruit hors service. Le sujet place le casque sur la tête, en mode « passif » uniquement (fig 6).

- 4) mesure du spectre de puissance du bruit sous le casque, réduction active de bruit en service. Le sujet toujours équipé du casque met en service la réduction active de bruit.

fig 6 *Mesure du niveau sonore sous le casque en mode passif ou actif*



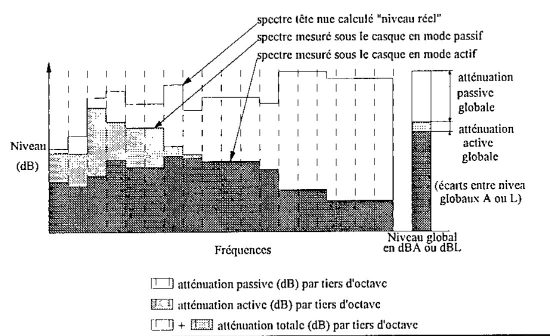
- 5) calcul de l'atténuation passive résultante. Par soustraction du spectre obtenu au 3), nous obtenons l'atténuation passive apportée par le protecteur auditif. L'atténuation passive est donnée pour chaque tiers d'octave : aux incertitudes de mesure près, ces valeurs sont indépendantes de l'ambiance sonore (atténuation passive linéaire). En outre, pour résumer les performances d'atténuation, nous présentons les écarts

entre niveaux globaux « tête nue » et « sous la protection », pondérés A ou non. Pour un casque donné, ces valeurs dépendent du type de bruit (atténuations analogues pour les bruits « ALPHA JET » et « ROSE », plus faible pour le bruit « PUMA »), la contribution énergétique de chaque bande de fréquences à l'énergie totale étant bien évidemment différente pour chaque bruit (par exemple, importance faible des fréquences supérieures à 3000 Hz dans le bruit PUMA, contrairement aux deux autres ambiances).

- 6) calcul de l'atténuation active résultante. Par soustraction du spectre obtenu au 4), nous obtenons l'atténuation active apportée par le dispositif. Ainsi que nous l'avons déjà souligné les valeurs d'atténuation sont susceptibles de varier selon le type de bruit, le filtre actif ayant un comportement différent en fonction de la fréquence et l'intensité du bruit.
- 7) calcul de l'atténuation totale. Elle est égale à la somme des atténuations passive et active calculées précédemment.

La figure 7 résume la démarche de mesure et de calcul développée ci-dessus.

fig 7 *Calcul des atténuations passive, active et totale*



1.3 INTELLIGIBILITE

1.3.1 les méthodes subjectives

Dans le cadre de la caractérisation de casques de communication en milieu bruité, la seule donnée des performances d'atténuation reflétant la qualité de la protection auditive n'est pas suffisante. Un protecteur auditif réellement efficace doit également apporter un confort d'écoute de la voie phonie se traduisant par une intelligibilité élevée des messages reçus par l'utilisateur. Ces remarques sont d'autant plus justifiées dans le cas de casques utilisant des circuits électroniques supplémentaires par lesquels doit transiter la phonie. Une conception du dispositif de réduction active de bruit de mauvaise qualité peut provoquer des dégradations sur le signal de phonie, la situation extrême étant représentée par une phonie considérée comme un bruit à annuler...

L'évaluation de l'intelligibilité repose essentiellement sur la mise en oeuvre de tests subjectifs : un expérimentateur doit identifier de la parole émise au travers du système à évaluer. Le terme « parole » est volontairement utilisé pour être le plus général possible. En effet, les tests d'intelligibilité se différencient par la nature de la parole émise : il peut s'agir des lettres de l'alphabet, des chiffres, de mots, de phrases, de diverses associations consonne-voyelle, mais également de composantes élémentaires du langage telles que les phonèmes. Ces diverses catégories constituent des « niveaux ».

Parallèlement aux scores d'intelligibilité, la qualité de restitution peut être évaluée par des questionnaires ou des « échelles » subjectives (sensation générale, clarté, timbre, bruit de fond...). Ce type d'expérience s'applique généralement à des canaux de communication possédant une grande intelligibilité et pour lesquels des tests fondés sur des scores d'intelligibilité ne sont pas significatifs en raison d'un effet « plafond ».

Nous décrivons plus précisément quelques tests représentatifs depuis le niveau « segment de parole » jusqu'au niveau « phrase ».

Le test « rime » est un test fréquemment utilisé pour déterminer les scores d'intelligibilité des phonèmes. Il s'agit d'un test à choix forcé dans lequel, après la présentation auditive de chaque mot, l'auditeur doit choisir une réponse parmi plusieurs présentées visuellement. En général, celles-ci ne diffèrent que d'un phonème à une position particulière dans le mot. Par exemple, pour un test sur une plosive en début de mot, nous pourrions avoir : Bom, Dom, Gom, Tom, Kom. Le test « rime » est facile à mettre en oeuvre et ne nécessite pas trop d'entraînement pour les auditeurs. Il existe deux types de test « rime » : le test « rime modifié » (MRT pour *Modified Rhyme Test*) -test des consonnes et des voyelles- et le test « rime diagnostique » (DRT pour *Diagnostic Rhyme Test*) -test des consonnes initiales uniquement. Le test MRT offre six choix, le DRT deux. STEENEKEN 92 (fig 8) a montré que ce dernier test est moins sélectif : en raison du faible choix proposé, les auditeurs sont parfois contraints de donner une réponse différente de leur impression perceptive (le phonème perçu peut ne pas être inclus dans les deux possibilités de réponse).

Une approche plus générale est obtenue grâce à un test à réponse « ouverte », ou libre, tels que les tests sur mots monosyllabiques, signifiants ou non, de type Consonne-Voyelle-Consonne (CVC), ou plus rarement de type CV, VC, CCVC ou encore CVCC. En utilisant des mots sans signification et en ayant une totale liberté de réponse, l'auditeur peut répondre par n'importe quelle combinaison de phonèmes. Cette procédure nécessite un entraînement des auditeurs. Les résultats peuvent être présentés sous forme de scores par phonèmes et par mots mais aussi de matrices de confusion entre consonnes initiales, voyelles, et consonnes finales. Des matrices de confusion créées par des réponses en choix ouvert fournissent des informations utiles à l'amélioration des performances d'un dispositif. Pour des tests de ce type, il est conseillé d'inclure ces mots dans des « phrases porteuses ». Celles-ci engendrent des échos et des

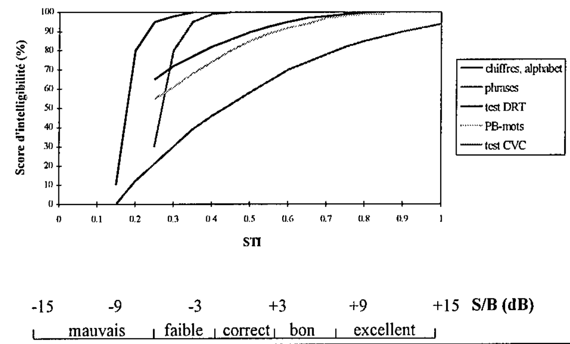
réverbérations représentatifs de distorsions temporelles. En outre, un réglage automatique de gain pourra être utilisé sur la phrase porteuse. L'importance de la phrase porteuse tient dans le fait qu'elle permet de stabiliser l'effort vocal du locuteur pendant la prononciation du mot-test. Le début de la phrase permet également d'annoncer l'émission du mot CVC.

L'intelligibilité des phrases est classiquement mesurée en demandant aux auditeurs d'estimer le pourcentage de mots correctement perçus grâce à une échelle 0-100 %. Cette méthode conduit à des résultats très variables selon les auditeurs. L'intelligibilité des phrases est très rapidement « saturée » à 100 %, même en présence d'un faible rapport signal à bruit.

L'attribution d'un « indice » de qualité est une méthode encore plus générale, utilisée pour évaluer l'acceptabilité des utilisateurs à l'égard d'un canal de transmission de parole. Des phrases de test normales ou bien une conversation libre permettent de recueillir l'impression des auditeurs. Ceux-ci doivent qualifier cette sensation sur une échelle subjective telle que : mauvais, faible, correct, bon, excellent. Là encore, le résultat appelé également « score d'opinion moyenne » présente une grande variabilité en fonction des auditeurs ; cet indice ne fournit pas de mesure absolue puisque les échelles ne sont pas calibrées, il est uniquement utilisé pour « classer » des équipements.

La figure 8 donne, pour cinq types de tests, le score d'intelligibilité en fonction du rapport signal à bruit de la parole entachée de bruit. Nous pouvons ainsi apprécier la dynamique effective de chaque test. La relation entre les scores d'intelligibilité et le rapport signal à bruit n'est valide que pour un bruit présentant des caractéristiques spectrales semblables à celles de la parole : le rapport signal à bruit est alors identique dans toutes les bandes de fréquences (un rapport signal à bruit de 0 dB signifie que parole et bruit ont la même densité spectrale). Comme nous pouvons le constater, les mots CVC (sans signification) sont discriminants sur une large dynamique, alors que des mots signifiants (généralement équilibrés phonétiquement : la distribution statistique des phonèmes est représentative du langage) présentent une dynamique plus faible. Les chiffres et l'alphabet montrent une saturation pour un rapport signal à bruit de -5 dB. Ce phénomène s'explique par le fait que les mots de tests sont en nombre très limité et que leur identification repose essentiellement sur la reconnaissance des voyelles. Le niveau moyen des voyelles est situé 5 dB au-dessus du niveau moyen des consonnes : elles sont donc plus « robustes » au bruit. Par contre, les distorsions non linéaires telles que la saturation auront un effet plus néfaste sur les voyelles que sur les consonnes. Les résultats de ces tests peuvent donc être faussés.

fig 8 *Comparaison de tests subjectifs d'intelligibilité (pour un bruit possédant un spectre fréquentiel similaire au spectre de la parole)STEENEKEN 92:*



1.3.2 STI : méthode objective par calcul de l'indice de transmission de la parole

Genèse et description (cf STEENEKEN 1992) :

L'idée fondamentale ayant conduit à la genèse de ce type de test consiste à prédire de façon objective l'intelligibilité, i.e. la qualité de transmission d'un système de communication vocale, par l'attribution d'un indice compris entre 0 et 1, appelé indice de transmission de la parole, et désigné sous l'abréviation de STI (Speech Transmission Index).

L'intérêt d'une telle évaluation objective réside essentiellement dans les deux points suivants :

- « standardisation » de la mesure autorisant une comparaison aisée entre systèmes. La donnée du STI peut être alors considérée comme une caractéristique du système équivalente à sa bande passante, son taux de distorsion, ou son rapport signal à bruit.
- rapidité de l'évaluation. Une mesure suffit à qualifier l'intelligibilité du système, ce qui constitue un gain de temps considérable par rapport aux évaluations subjectives.

Par contre, pour confirmer sa validité, cet indice doit correspondre (i.e. être fortement corrélé) à un pourcentage d'intelligibilité obtenu par des tests subjectifs. C'est pourquoi la mise au point de la méthodologie de mesure du STI a été effectuée grâce à une procédure exploitant de façon itérative mesures objectives et subjectives obtenues par test CVC (se reporter à la description de ce type de test ci-dessus). C'est pourquoi à toute valeur de STI est associée un pourcentage d'intelligibilité sur mot CVC (voir également la figure ci-dessus).

Le principe de la méthode consiste à émettre un signal de « test » dans le système à évaluer, à analyser le signal récupéré en sortie puis de calculer l'indice en fonction des déformations et dégradations subies dans différentes bandes de fréquences. Nous détaillons ci-dessous l'analyse effectuée et le calcul du STI.

Le signal de test est un « bruit » dont le spectre est égal à celui de la parole. Il est obtenu de la manière suivante : pour chaque bande de fréquences centrée sur

f_0, f_0 couvrant sept octaves de 125 à 8000 Hz, on réalise une modulation sinusoïdale d'intensité spectrale (modulation d'amplitude par $\sqrt{1+\cos(2.\pi.f_m.t)}$).

L'analyse du signal de test s'effectue pour chaque octave de la manière suivante : filtrage passe-bande centré sur f_0, puis détection d'enveloppe conduisant par analyse de Fourier aux indices de modulation ($m_{f,k}$) pour chaque bande k de fréquences centrée sur f_0, à la fréquence de modulation f, et à l'intensité (I_k) du signal reçu dans chacune des bandes. Les fréquences de modulation f utilisées pour l'analyse du signal sont réparties par tiers d'octave entre 0,63 et 12,5 Hz soit 14 fréquences.

Outre le phénomène de masquage dû au bruit perturbant le canal de communication, il faut tenir compte d'un phénomène additionnel de masquage auditif : du fait de la physiologie, les sons graves masquent les sons aigus. Cet effet est modélisé sous la forme d'un bruit de masquage imaginaire qui conduit à une diminution du rapport signal à bruit effectif et à une réduction de l'indice de modulation. Dans l'approche propre au calcul du STI, cet effet ne dépend ni de la bande de fréquences considérée ni de son niveau, et il décroît de 35 dB par octave, ce qui fournit un facteur de « masquage auditif » de fma = 0,000316. Ainsi l'intensité du signal dans la bande k dû à l'influence de la bande k-1 est : $I_{ma,k} = I_{k-1} * f_{ma}$. (seul l'effet de masquage attribué à la bande inférieure la plus proche étant significatif). La prise en compte de cet effet conduit à un indice de modulation corrigé $m'_{k,f}$:

$$m'_{k,f} = m_{k,f} \cdot \frac{1}{I_k + I_{ma,k}}$$

Le rapport signal à bruit effectif dans la bande de fréquences k et la fréquence de modulation f devient alors en dB :

$$(S/B)_{k,f} = 10.\log\left(\frac{m'_{k,f}}{1-m'_{k,f}}\right)$$

Selon le concept du STI, un rapport signal à bruit compris entre -15 et +15 dB est associé à une contribution à l'intelligibilité globale comprise entre 0 et 1. Par conséquent, le rapport signal à bruit effectif est converti en un indice de transmission $TI_{k,f}$, spécifique à la bande k et à la fréquence de modulation f par la formule :

$$TI_{k,f} = \frac{(S/B)_{k,f} + 15}{30}, \text{ où } 0 \le TI_{k,f} \le 1$$

Pour chaque octave k, la moyenne des 14 indices de transmission fournit un indice de transfert de modulation MTI_k :

$$MTI_k = \frac{1}{14} \cdot \sum_{f=1}^{14} TI_{k,f}$$

Finalement, l'indice STI est obtenu comme la somme pondérée des différents indices de transfert de modulation pour l'ensemble des sept octaves :

$$STI = \sum_{k=1}^{7} \alpha_k . MTI_k \text{ avec } \sum_{k=1}^{7} \alpha_k = 1$$

où les α_k représentent les facteurs de pondération d'octave. Suite à des études plus poussées, cette formule a été modifiée afin de compenser la contribution des bandes fréquentielles adjacentes. Cette contribution est traduite par l'ajout de coefficients de « redondance » β_k ce qui donne :

$$STI = \sum_{k=1}^{7} \alpha_k . MTI_k - \sum_{k=1}^{6} \beta_k . \sqrt{MTI_k - MTI_{k+1}}$$

$$\text{avec } \sum_{k=1}^{7} \alpha_k - \sum_{k=1}^{6} \beta_k = 1$$

La détermination des coefficients α_k et β_k optimaux pour les voix masculine et féminine ainsi que pour les différents groupes de phonèmes résulte d'une procédure itérative comparant STI et tests subjectifs.

Remarque à propos du RASTI (Room Acoustical Speech Transmission Index) :

C'est une forme simplifiée du STI qui a été validée au niveau international (norme IEC 268). L'analyse est limitée aux bandes d'octaves 500Hz et 2000Hz. Les distorsions non linéaires ne sont pas prises en compte, le bruit de fond doit être stationnaire et ne pas contenir de raies de forte intensité, enfin la bande passante ne doit pas être limitée. Le RASTI est utilisé en acoustique des salles.

Mise en oeuvre pratique:

La mise en oeuvre du test STI est aujourd'hui totalement informatisée : le signal de test numérisé est stocké sur le disque dur d'un micro-ordinateur de type PC ; la restitution sonore (dans les écouteurs du casque, en ce qui nous concerne) est effectuée par l'intermédiaire d'une carte de conversion numérique-analogique ; le signal de test ainsi émis est capturé par un microphone (dans notre cas, le microphone miniature de la méthode MIRE), puis numérisé et enfin rangé en mémoire RAM de l'ordinateur.

Un logiciel dédié assure la gestion en temps réel de la restitution et de l'enregistrement simultanés. Un second logiciel analyse le signal en temps différé et détermine le STI correspondant. La durée totale du test (restitution-enregistrement + analyse) est de 30 secondes environ (version « STITEL » du signal de test, destinée aux systèmes de télécommunication).

Il existe en effet différents signaux de test selon les systèmes à caractériser. Celui utilisé pour nos essais est le STITEL, analyse réduite peu robuste à la distorsion non linéaire ou temporelle (7 octaves, 7 fréquences de modulation associées aux octaves ; durée du signal :

15s). Les différentes fréquences de modulation f_m nécessaires à la génération du signal sont associées aux octaves de la manière suivante :

f_0 (Hz)	125	250	500	1000	2000	4000	8000
f_m (Hz)	1,12	11,33	0,71	2,83	6,97	1,78	4,53

L'exhaustivité de la méthode explique son utilisation dans de nombreux domaines dans le monde entier, autant dans le milieu civil que les armées (Etats-Unis, Hollande, Belgique, Canada...). Certains constructeurs/clients de matériel de télécommunications imposent la caractérisation des systèmes par cette méthode. Dans ce cadre, un indice STI minimal de **0,6** est nécessaire pour qualifier un équipement évalué en laboratoire (meilleur cas) ; un indice minimal de **0,35** doit être garanti en fonctionnement nominal (utilisation courante). En deçà de ce second seuil théorique, l'intelligibilité est dégradée de telle sorte que l'utilisateur d'un système de communication (téléphone, radio, casque de communication...) soit obligé de demander une confirmation ou une répétition à son interlocuteur.

Ces deux valeurs sont utiles pour comparer les casques en mode passif et actif, dans les différentes ambiances.

Afin d'obtenir des mesures totalement objectives, il est nécessaire d'ajuster le volume sonore du signal de test, pour chaque sujet, pour chaque casque. Il est en effet évident que toute variation relativement conséquente du volume d'écoute influe considérablement sur l'intelligibilité. Nous avons donc décidé de restituer le signal de test à un niveau de 83 dBA, ce qui correspond à l'écoute de la parole « continue » (texte ponctué) à un niveau de 85 dBA. Cette valeur autorise théoriquement une écoute quotidienne de huit heures sans dégradation de l'audition de l'opérateur. Il est intéressant de noter que les concepteurs du test STI préconisent un niveau de 70-75 dBA pour le signal de test : dans le cadre de nos expérimentations, un tel choix aurait conduit à des résultats désastreux (et donc inexploitables) pour certains des casques évalués, en particulier pour le casque pilote actuel !

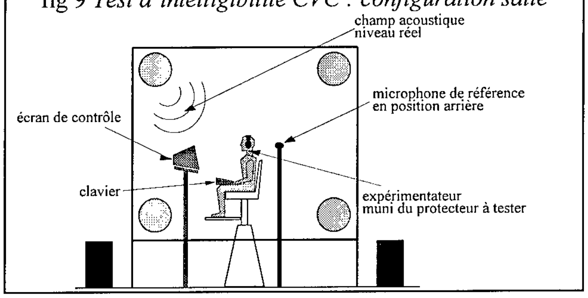
En pratique, le sujet est équipé du microphone intra-auriculaire et du casque à évaluer ; un atténuateur calibré permet de régler le volume sonore mesuré sous le casque pendant l'émission du signal de test. Cette opération est bien évidemment nécessaire pour chaque sujet (différence d'appui des coques sur la tête, cavités acoustiques « oreille-écouteur » dissemblables), pour chaque casque (sensibilités différentes des écouteurs), et pour chaque mode de fonctionnement (actif ou passif) : ce dernier point est d'importance capitale car la mise en service de la réduction active de bruit peut modifier considérablement la voie phonie. Du fait d'une variation de l'impédance d'entrée, certains casques amplifient la phonie (jusqu'à 10-12 dB) lorsqu'ils passent en mode actif : il en résulte une augmentation artificielle de l'intelligibilité obtenue au détriment des oreilles de l'auditeur.

1.3.3 comparaison CVC / STI

Le but de cette comparaison n'est pas de valider rigoureusement l'usage du STI pour la langue française. Une telle étude nécessiterait beaucoup de temps et de sujets puisqu'il conviendrait de tester pour divers types de bruit et de distorsions du canal de communication. Il s'agit simplement de se placer dans deux ambiances sonores réalistes (Alphajet et Puma) en se limitant à deux casques, à savoir le GUENEAU 458 utilisé en aéronautique militaire, et un prototype ANR monté dans le même type de casque. Notons qu'une validation multi-langues du RASTI (y compris langue française) a été effectuée par HOUTGAST (1984). L'approche par test CVC permet aux expérimentateurs de rester en contact avec la réalité des communications vocales. En outre le STI fournissant une prédiction d'intelligibilité sur score CVC, il est intéressant de pouvoir effectuer la comparaison entre prédiction et score CVC.

Le type de test subjectif employé présente des réponses à choix libres et discrimination importante sur une large plage de rapport signal à bruit. L'auditeur doit reconnaître un mot Consonne Voyelle Consonne inséré dans une phrase porteuse. Les mots CVC sont construits à partir des consonnes initiales, des voyelles et des consonnes finales, apparaissant dans la langue parlée. 19 consonnes initiales et finales ainsi que 11 voyelles ont été retenues. Ces mots CVC tirés au hasard sont placés dans 57 phrases porteuses différentes. Ces phrases et les CVC sont enregistrés et joués au moyen d'un support informatique. Pratiquement, le sujet écoute chaque parole de test et saisit au clavier les phonèmes reconnues du CVC (fig 9). Le logiciel gère le test et comptabilise les erreurs.

fig 9 *Test d'intelligibilité CVC : configuration salle*



Notre analyse se voulant relativement synthétique, elle portera essentiellement sur les pourcentages d'intelligibilité (en médiane et non en moyenne pour limiter les effets systématiques d'apprentissage, les pertes temporaires d'attention et les erreurs de réglage du volume) sur mots CVC. En ambiance ALPHA JET, 10 tests ont été effectués (5 auditeurs, 2 locuteurs) pour chaque configuration de casque ; en ambiance PUMA également, excepté pour le casque GUENEAU 458, pour lequel 8 tests seulement ont été réalisés. Ces résultats sont rassemblés dans les tableaux ci-après ainsi que les indices STI et les prédictions associées d'intelligibilité sur mot CVC obtenus dans les mêmes conditions expérimentales.

bruit	Alphajet		
casques	% intelligibilité cvc	STI	% prédiction cvc
GUENEAU 458	41,5	0,30	31
PROTO 458 ANR OFF	75	0,69	76,5
PROTO 458 ANR ON	82	0,76	81,5

bruit	PUMA		
casques	% intelligibilité cvc	STI	% prédiction cv
GUENEAU 458	32	0,34	37,5
PROTO 458 ANR OFF	76	0,69	76,5
PROTO 458 ANR ON	76,5	0,77	82

Il convient donc de noter la cohérence correcte entre les scores prédits par la mesure du STI et les résultats des tests CVC : intelligibilité du casque GUENEAU 458 de l'ordre de 35-40 %, du casque PROTO de l'ordre de 75-80 % avec réduction active de bruit. Les correspondances établies entre résultats objectifs et subjectifs sont encourageantes pour la confirmation de leur validation réciproque et nous autorise à utiliser le test objectif STI.

2. MOYENS

2.1 ESSAIS EN SITUATION LIMITE

Dans cette approche, deux types d'expérimentation ont été retenus :

2.1.1 bruit rose 120dBA

Les essais sur tête artificielle NEUMAN sont réalisés dans une enceinte réverbérante de très faibles dimensions, « alimentée » en bruit rose non égalisé (la mise en place de la tête modifiant considérablement le champ acoustique dans une zone de si petites dimensions, l'égalisation ne représente rien de significatif).

Dispositif expérimental:

Les faces avant de quatre enceintes BOSE 802 forment les quatre faces verticales d'un cube de 50 cm de côté. La tête artificielle est placée au centre de ce cube sur un pied servant également de support au microphone de référence, lors de la mesure préliminaire du niveau en l'absence de tête artificielle. Une plaque de PVC recouvre les enceintes afin d'augmenter le niveau de bruit généré, par réverbération. Lors des expériences, un microphone de référence est placé dans l'un des angles du dispositif pour contrôler le niveau. Le niveau maximal atteint est de 120 dBA-125 dBLin.

2.1.2 sinus basse fréquence fort niveaux (<135dBLin)

Les essais sur tête artificielle NEUMAN se déroulent dans un caisson clos de très faibles dimensions, « alimenté » par un signal sinusoïdal pur à basse fréquence (40 Hz).

Dispositif expérimental:

Deux caissons de basse (BOSE 302) sont disposés face à face à 50 cm de distance. Des plaques (bois et PVC) sont utilisées pour fermer l'enceinte ainsi constituée (deux côtés et plafond). Le microphone est placé au centre du dispositif pour les mesures de référence en l'absence de la tête artificielle. Le niveau maximal atteint est de 135 dB à 40 Hz. L'homogénéité du champ à l'intérieur du caisson fermé et la faible perturbation due à la présence de la tête ont été vérifiées. La distorsion harmonique intrinsèque de l'installation a été contrôlée pendant les essais.

2.2 SALLE DE RESTITUTION D'AMBIANCE SONORE

2.2.1 description de la salle

Le champ acoustique diffus a été obtenu au sein d'une chambre semi-anéchoïque. Sur une structure cubique (arête 4,30m) sont installées huit enceintes acoustiques BOSE 802 médium-aigu. Les sons graves (peu directifs) sont produits par deux caissons BOSE 402 posés au sol sur des supports inclinés (fig 10, Photo 2). La diffusion est obtenue en insérant dans la chaîne de puissance des retards numériques programmables (fig 11). L'expérimentateur est installé sur un siège surélevé de telle sorte que sa tête se trouve dans la zone de champ diffus. Le sujet se positionne par rapport à des repères video de telle sorte que sa tête soit toujours au centre du champs diffus sans aucune contrainte mécanique.

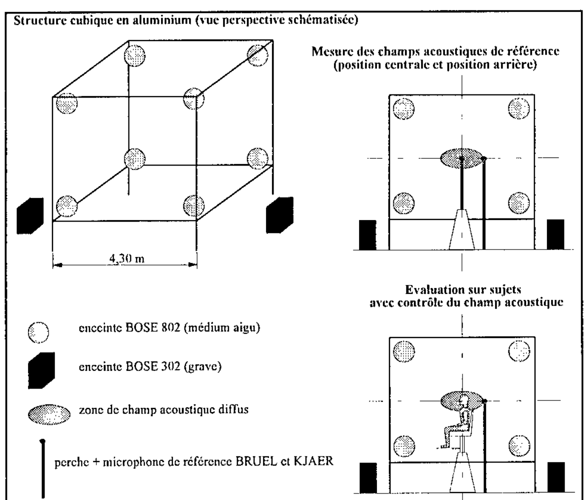


fig 10 *Salle de restitution d'ambiance sonore*

2.2.2 homogénéité du champ

Le champ acoustique restitué doit être diffus : ses caractéristiques (spectre, niveau) doivent être identiques dans un volume entourant l'expérimentateur. Ainsi, les mesures de niveau sonore ne sont théoriquement pas influencées par l'attitude du sujet (l'orientation de sa tête en particulier).

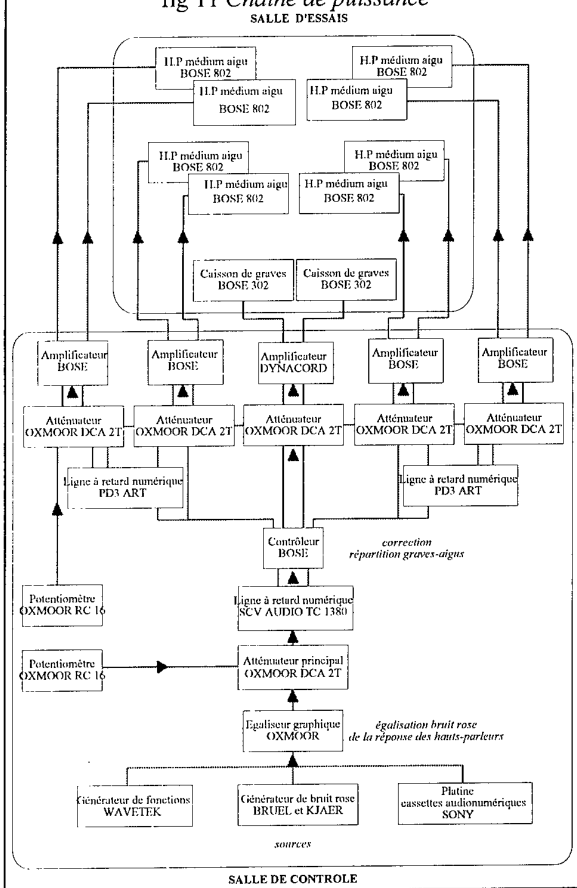
L'ensemble de ce dispositif permet de restituer de manière parfaite des ambiances réalistes en spectre et en niveau. Seules les fréquences en dessous de 50 Hz et

au delà de 12500 Hz ne sont pas représentatives, compte tenu de la réponse des haut-parleurs. Une étude sur le champ acoustique a été réalisée. Ce champ diffus est conforme intégralement à la norme ISO 4869 et conforme à la norme ANSI S12.6 excepté pour une symétrie concernant le tiers d'octave 8kHz ou le dépassement est de 0,3dB par rapport à la tolérance préconisée.

2.2.3 invariance temporelle du champ sonore

Après quelques minutes de chauffe à fort niveau, en début de séance, l'invariance est assurée pour chaque tiers d'octave. En outre, un second microphone de référence est installé à la même hauteur mais 500 mm en arrière. Une mesure de référence (sans sujet) pour chacune des ambiances fournit le niveau et le spectre résultant à la position « arrière ». Ce microphone reste à demeure pendant toute la durée des expérimentations et permet un contrôle permanent du niveau et du spectre restitués en salle lors des essais sur sujets.

fig 11 *Chaîne de puissance*

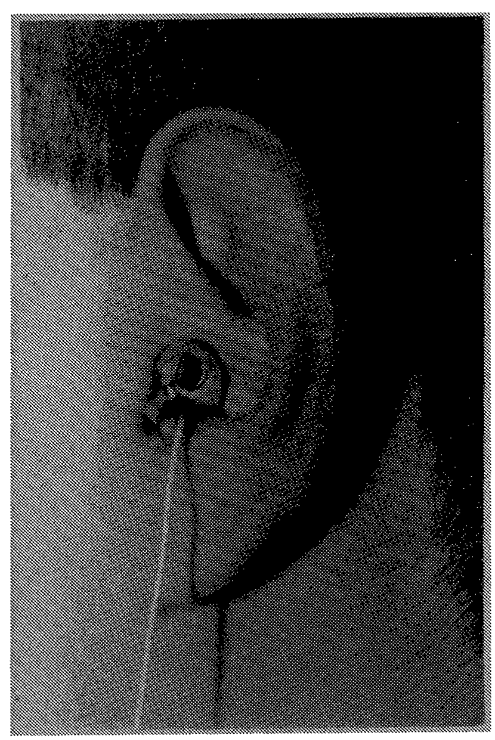


2.3 CHAINE DE MESURE MICROPHONIQUE

Le microphone miniature inséré dans l'oreille est un SENNHEISER, MKE-4.211(diamètre 5mm) à électret. Sa réponse en fréquence est plate dans la bande 20 Hz-20 kHz, avec une dynamique importante. Les mesures de STI sont également effectuées avec ce microphone (fig 12).

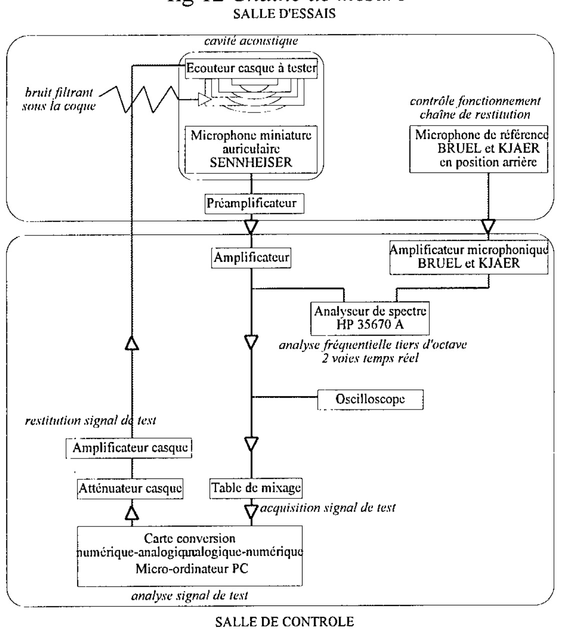
Le maintien du bouchon à l'entrée du conduit auditif est traditionnellement assuré par un support en fil d'acier plastifié semi-rigide entourant le pavillon de l'oreille. Si ce dispositif offre l'avantage de conserver toutes les facultés auditives du sujet (des tests subjectifs et objectifs peuvent ainsi être menés en parallèle), il n'a pas paru totalement satisfaisant du point de vue de la reproductibilité du positionnement du microphone et donc, des mesures. En outre, l'expérimentation comprend l'évaluation de casques intégraux dont le positionnement sur la tête est délicat sans déplacer le microphone. Ce dernier a donc été enveloppé de mousse souple empruntée à un bouchon d'oreille BILSOM « FORM » puis inséré dans un bouchon BILSOM « QUIETZONE ». Les faibles dimensions de l'assemblage permettent son insertion parfaite dans le conduit auditif. Le microphone est alors totalement immobile par rapport au conduit (Photo 1); la membrane sensible est située, selon les sujets, soit à l'entrée du conduit, soit plus proche du tympan, à moins de 10 mm de ce dernier (longueur du bouchon BILSOM).

Photo1 : *Méthode MIRE*



Les mesures effectuées ont démontré la qualité de ce dispositif (reproductibilité importante pour un même sujet même après retrait et nouvelle mise en place du bouchon, faible écart-type entre sujets) et sa facilité d'utilisation (peu de gêne, mise en place aisée de tous les casques).

fig 12 *Chaîne de mesure*



2.4 CONDITIONS DE MESURE

2.4.1 Ambiances sonores

Compte tenu de la particularité de cette étude, orientée vers l'application de la réduction active de bruit à l'aéronautique, deux environnements sonores rencontrés à bord d'aéronefs actuellement en service, ainsi qu'une ambiance « standard » ont été retenus (fig 13).

- ambiance de type ALPHA JET : nous utilisons un enregistrement original effectué en vol, à bord d'un avion à réaction de type ALPHA JET, grâce à des microphones placés sur le casque ou sur le masque du pilote. Une fraction de cet enregistrement (5 à 10 secondes) a été sélectionnée puis bouclée sur elle-même grâce à un logiciel de mixage numérique afin de réaliser un nouvel enregistrement pendant toute la durée duquel le bruit reste stable. La calibration de l'enregistrement original (au début de la bande, a été enregistré le signal émis par un piston-phone) permet de connaître le niveau réel et de le restituer en salle, soit 101,9 dBA,
- ambiance de type PUMA (id. avec un enregistrement à bord d'un hélicoptère PUMA). Le niveau obtenu en salle est de 102,0 dBA,
- ambiance BRUIT ROSE : à partir d'un générateur de bruit rose BRUEL et KJAER, nous restituons un bruit dont le spectre est plat à ± 1dB dans la plage de fréquence 50-12500 Hz. Le niveau global obtenu est de 106,8 dBA. Ce type de bruit est aisément reproductible une fois son niveau déterminé ; il est classiquement utilisé comme référence dans les évaluations acoustiques.

Photo 2 :*salle d'essais*

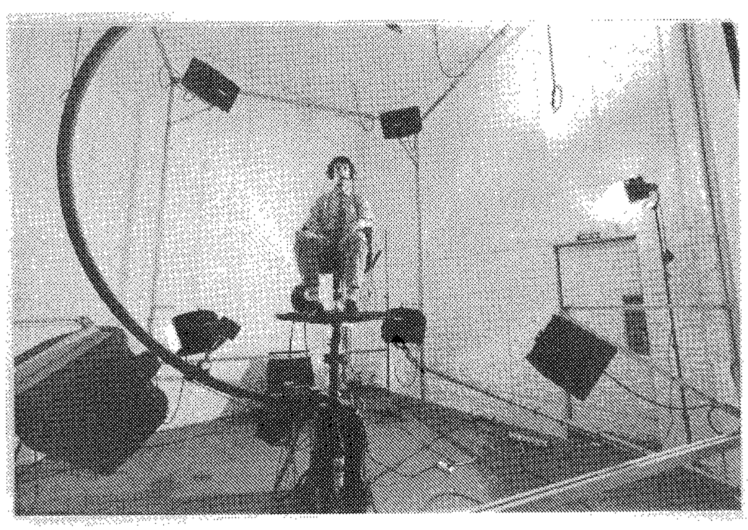
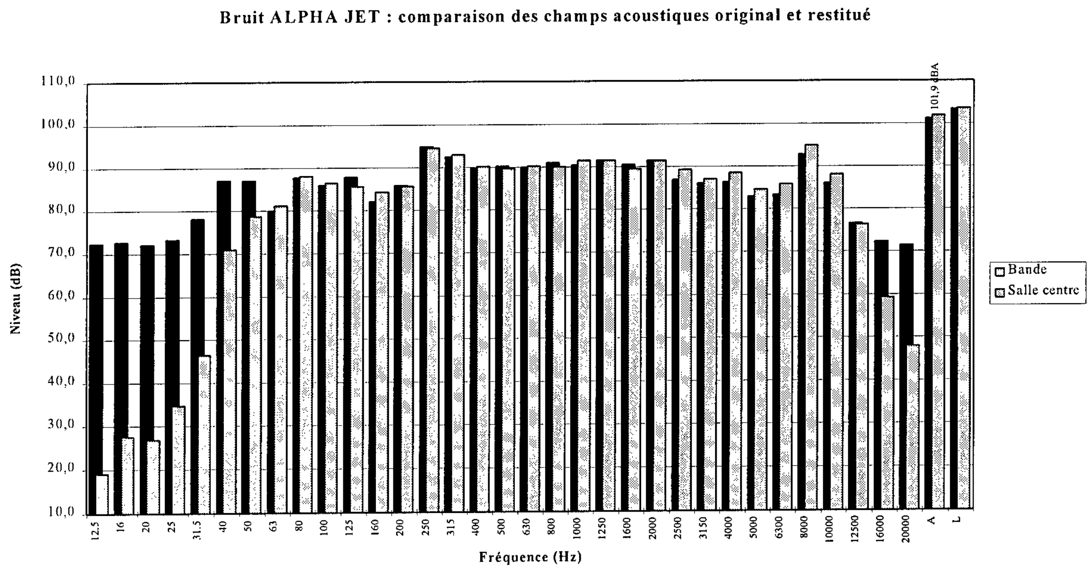
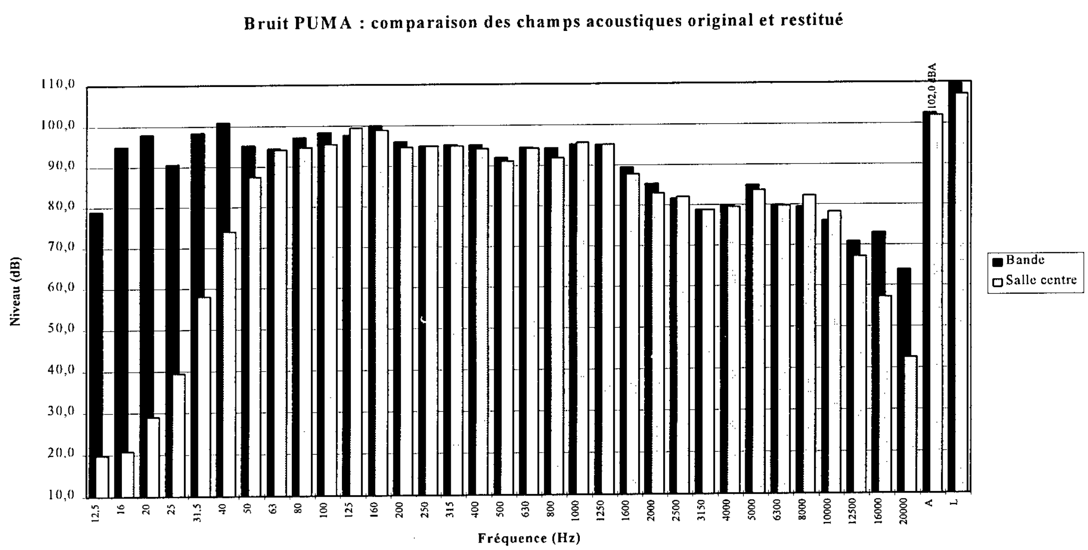
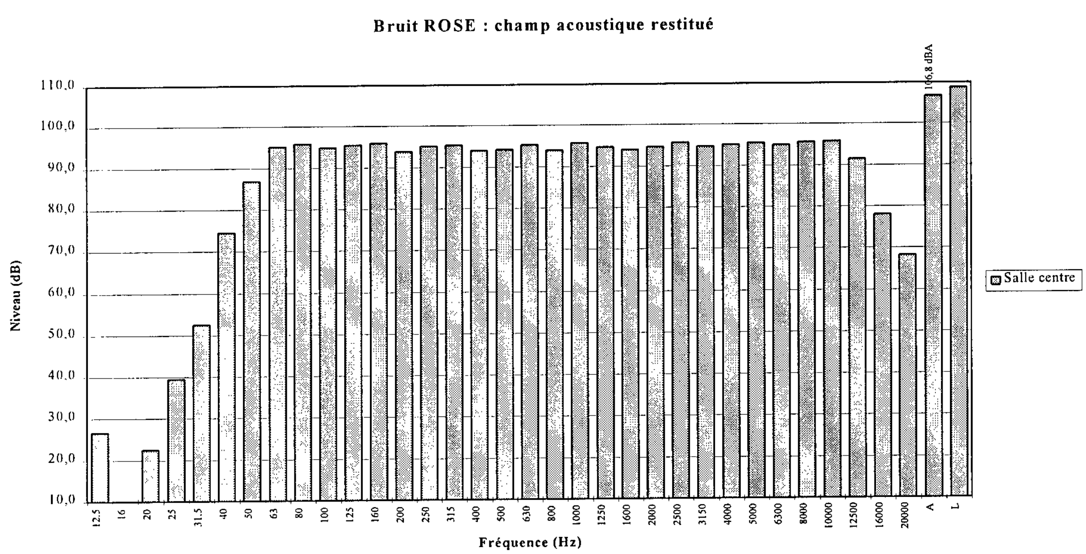


fig 13 *Comparaison des spectres en tiers d'octaves des ambiances originales et restituées*







2.4.2 sujets

Cinq sujets masculins ont participé à l'étude. Leurs caractéristiques morphologiques crâniennes sont à priori normales et leurs coiffures adaptées au port des casques intégraux. Tous les équipements ont été ajustés au mieux vis à vis du confort et de l'étanchéité au bruit.

sujets	DS	JD	GR	LP	PB
largeur de tête en cm	15,2	15,6	17,6	15,5	17,2

2.4.3 casques

En Europe et aux Etats-Unis de nombreuses sociétés commercialisent des casques antibruit (essentiellement de type serre-tête) utilisant la réduction active de bruit. Ils sont destinés à l'aviation civile légère de transport ou de convoyage ainsi qu'aux hélicoptères civils. Dernièrement est apparue dans les forces armées notamment américaines, l'utilisation opérationnelle de tels casques en version militarisée. L'évaluation des performances de la protection auditive et de l'intelligibilité combinant atténuations passive et active a été effectuée à partir d'un échantillonnage représentatif de casques antibruit disponibles commercialement (1994) donc principalement civils. L'ensemble des huit systèmes ANR essayés est composé de :

- 3 casques serre-tête de fabrication américaine, le BOSE, le TELEX ANR et le TELEX ANR 4000,

- 1 système ANR de fabrication anglaise sous forme de coques et montées sur arceau, le HELMET,

- 1 casque serre-tête de fabrication allemande le SENNHEISER HD EC 200,

- 3 système ANR néerlandais, le casque serre-tête TNO, et 2 casques intégraux prototypes :
  - un casque TNO-GUENEAU 458 modifié dont la coque protectrice a été remodelée partiellement pour intégrer les coques auditives du casque TNO,
  - un casque PROTO 458 ANR. Il s'agit d'un GUENEAU 458 équipé d'écouteurs redessinés et qualifié bon de vol.

A cette liste vient s'ajouter le casque intégral français GUENEAU 458, en dotation dans l'Armée de l'Air, permettant d'évaluer l'apport de la réduction active de bruit.

La plupart des système ANR fonctionne sur batteries mais pour des raisons de commodité et de fiabilité lors des expérimentations, nous avons modifié légèrement le câblage électrique d'alimentation pour pouvoir connecter ces casques à une alimentation continue, à tension réglable et stabilisée. D'autre part nous avons mesuré la force d'appui de chacun de ces casques dans les conditions utilisées par les sujets. Cette force d'appui reste constante quels que soient les sujets.

modèle	**code**	**alimentation**	**pression tête en daN**
BOSE Aviation	BOS	0-20V	0,9
GUENAU 458	GUE	aucune	non mesuré*
HELMETT	HEL	0-18V	0,9
PROTO 458 ANR	PRO	0-20V	non mesuré*
SENNHEISER	SEN	0-20V	0,9
TELEX ANR 4000	TE4	0-9V	0,9
TELEX ANR	TEL	0-9V X2	1,2
TNO GUENAU 458	TNG	0-20V	non mesuré*
TNO Peltor	TNO	0-20V	0,9

non mesuré*: casque intégral

Les casques sont connectés sur la sortie « casque » d'un amplificateur de puissance STUDER_A68. L'impédance de sortie correspondante est de 130 Ohms. Une étude électroacoustique sur l'influence de l'impédance de sortie du générateur a révélé que la répercussion sur la mesure d'intelligibilité est négligeable pour tous les casques, à l'exception du BOSE en mode OFF pour lequel elle est minime. Pour ce dernier modèle, un test en ambiance bruit rose (85dBA) montre que le STI varie de 0,79 à 0,86 pour des impédances de sortie variant de 0 à 600 Ohm.

3 RESULTATS

3.1 RESULTATS EN SITUATION LIMITE

3.1.1 essais avec fuites acoustiques, chocs, et coupure électrique

Le tableau ci-dessous regroupe les résultats en ambiance silencieuse et bruyante.

essais:	fuites acoustiques	retrait	chocs	coupure alimentation
effet :	instabilité	son regénéré	désactivation spontanée de l'ANR	bruits de commutation
casques:				
BOS	**non**	**non**	**oui**	**faible**
HEL	**oui**	**non**	**non**	**désagréable**
PRO	**non**	**non**	**non**	**important**
SEN	**non**	**non**	**oui**	**important**
TE4	**non**	**non**	**non**	**aucun**
TEL	**non**	**oui**	**non**	**désagréable**
TNG	**non**	**non**	**non**	**important**
TNO	**non**	**non**	**non**	**important**

3.1.2 Essai bruit rose 120dBA

Dans le tableau ci-dessous la rubrique « largeur de bande » correspond à la bande de fréquence pour laquelle une atténuation active est observée. Cette atténuation figure dans la rubrique « att ». Les valeurs indiquées correspondent au minimum et maximum observés dans les tiers d'octave.

observations des systèmes ANR sous bruit rose 120dBA				
	largeur de bande en Hz	att en dB	regénération de bruit	att globale en dBA
casques:				
BOS	**125-400**	**1 à 10**	**3dB max de 50 à 100Hz**	**5**
HEL	**125-250**	**0 à 5**	**15dB max de 315 et 3150Hz de**	**<1 non significatif**
PRO	**50-630**	**1 à 28**	**<3 dB de 630 à 2000Hz (visières fermées)**	**10**
SEN	**pas**	**d'effet**	**ANR**	**constaté**
TE4	**100-400**	**0 à 12**	**3dBmax de 630 à 2000Hz**	**9**
TEL	**50-400**	**5 à 20**	**<3dB de 630 à 2000Hz**	**10**
TNG	**50-630**	**2 à 12**	**aucun**	**7**
TNO	**50-630**	**2 à 10**	**2dBmax de 2500 à 3150Hz**	**5**

3.1.3 Essai sinus 40Hz fort niveau

sinus 40Hz	dépassement du seuil 20 % de distorsion au niveau sonore de:
casques:	(dB)
BOS	107
HEL	instabilité
PRO	107
SEN	120
TE4	122
TEL	116
TNG	133
TNO	122

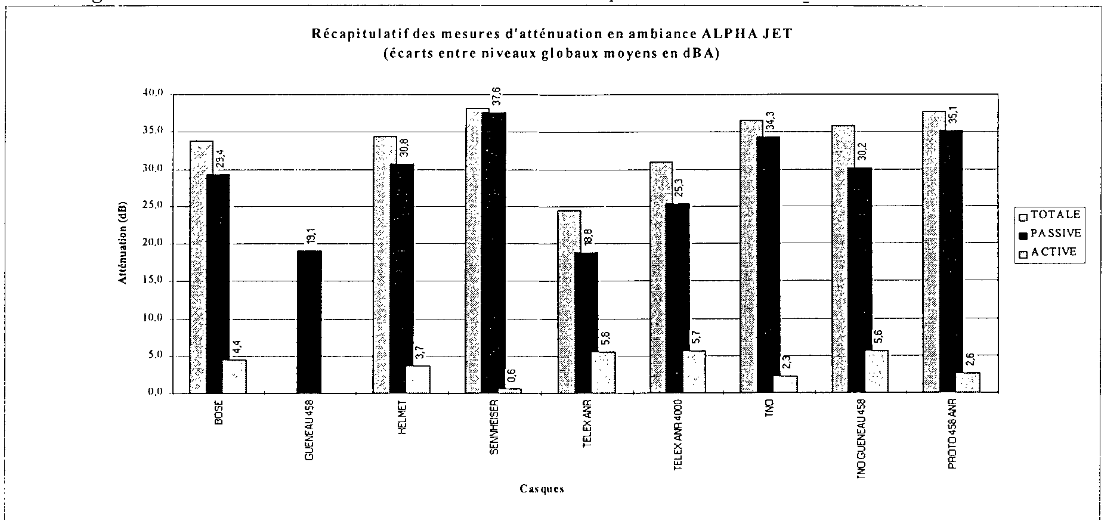
3.2 MESURE D'ATTENUATION

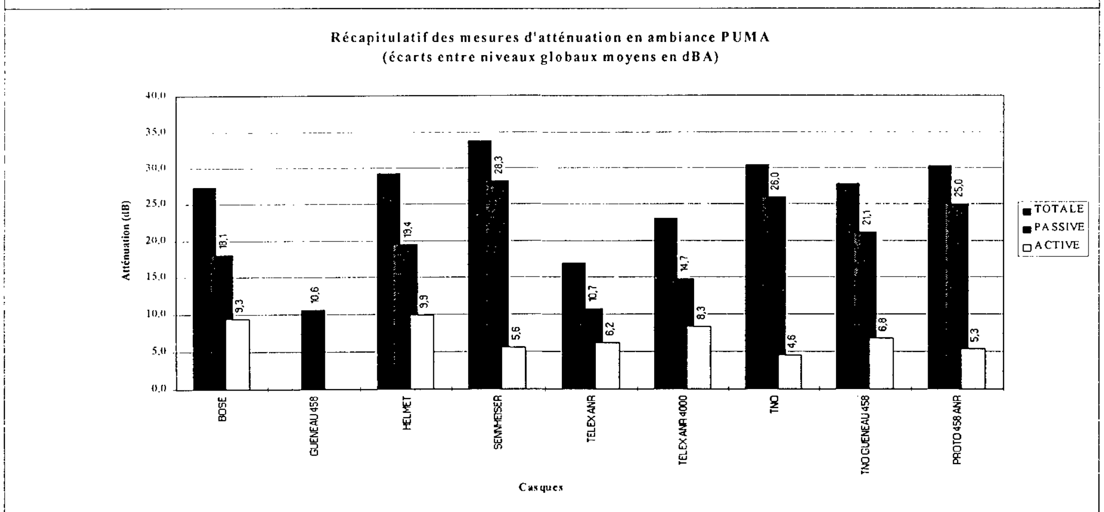
BRUIT ALPHAJET						
MOYENNE ATTE	PERFORMANCES (dBA)			**ATTENUATION (dB)**		
	REEL	OFF	ON	PASSIVE	ACTIVE	TOTALE
BOSE	109,3	80,0	75,5	**29,4**	**4,4**	**33,8**
GUENEAU 458	109,3	90,2		**19,1**		
HELMET	109,3	78,6	74,9	**30,8**	**3,7**	**34,5**
SENNHEISER	109,3	71,7	71,1	**37,6**	**0,6**	**38,3**
TELEX ANR	109,3	90,5	84,9	**18,8**	**5,6**	**24,4**
TELEX ANR 4000	109,3	84,0	78,3	**25,3**	**5,7**	**31,0**
TNO	109,3	75,0	72,8	**34,3**	**2,3**	**36,6**
TNO GUENEAU 458	109,3	79,2	73,5	**30,2**	**5,6**	**35,8**
PROTO 458 ANR	109,3	74,2	71,6	**35,1**	**2,6**	**37,7**

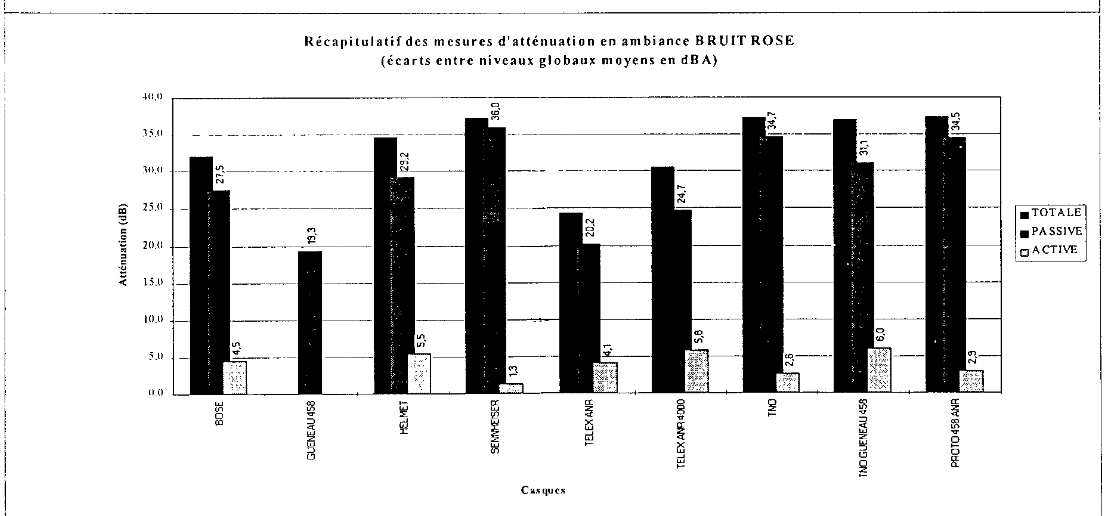
BRUIT PUMA						
MOYENNE ATTE	PERFORMANCES (dBA)			**ATTENUATION (dB)**		
	REEL	OFF	ON	PASSIVE	ACTIVE	TOTALE
BOSE	104,3	86,2	76,9	**18,1**	**9,3**	**27,4**
GUENEAU 458	104,3	93,7		**10,6**		
HELMET	104,3	84,9	75,0	**19,4**	**9,9**	**29,3**
SENNHEISER	104,3	76,0	70,4	**28,3**	**5,6**	**33,9**
TELEX ANR	104,3	93,5	87,4	**10,7**	**6,2**	**16,9**
TELEX ANR 4000	104,3	89,5	81,3	**14,7**	**8,3**	**23,0**
TNO	104,3	78,3	73,8	**26,0**	**4,6**	**30,5**
TNO GUENEAU 458	104,3	83,2	76,3	**21,1**	**6,8**	**27,9**
PROTO 458 ANR	104,3	79,3	73,9	**25,0**	**5,3**	**30,3**

BRUIT ROSE						
MOYENNE ATTE	PERFORMANCES (dBA)			**ATTENUATION (dB)**		
	REEL	OFF	ON	PASSIVE	ACTIVE	TOTALE
BOSE	113,6	86,0	81,5	**27,5**	**4,5**	**32,1**
GUENEAU 458	113,6	94,3		**19,3**		
HELMET	113,6	84,4	78,9	**29,2**	**5,5**	**34,7**
SENNHEISER	113,6	77,6	76,3	**36,0**	**1,3**	**37,3**
TELEX ANR	113,6	93,4	89,2	**20,2**	**4,1**	**24,3**
TELEX ANR 4000	113,6	88,9	83,1	**24,7**	**5,8**	**30,4**
TNO	113,6	78,9	76,3	**34,7**	**2,6**	**37,3**
TNO GUENEAU 458	113,6	82,5	76,5	**31,1**	**6,0**	**37,1**
PROTO 458 ANR	113,6	79,1	76,2	**34,5**	**2,9**	**37,4**

barre gauche : atténuation totale *barre centrale : atténuation passive* *barre droite : atténuation active*







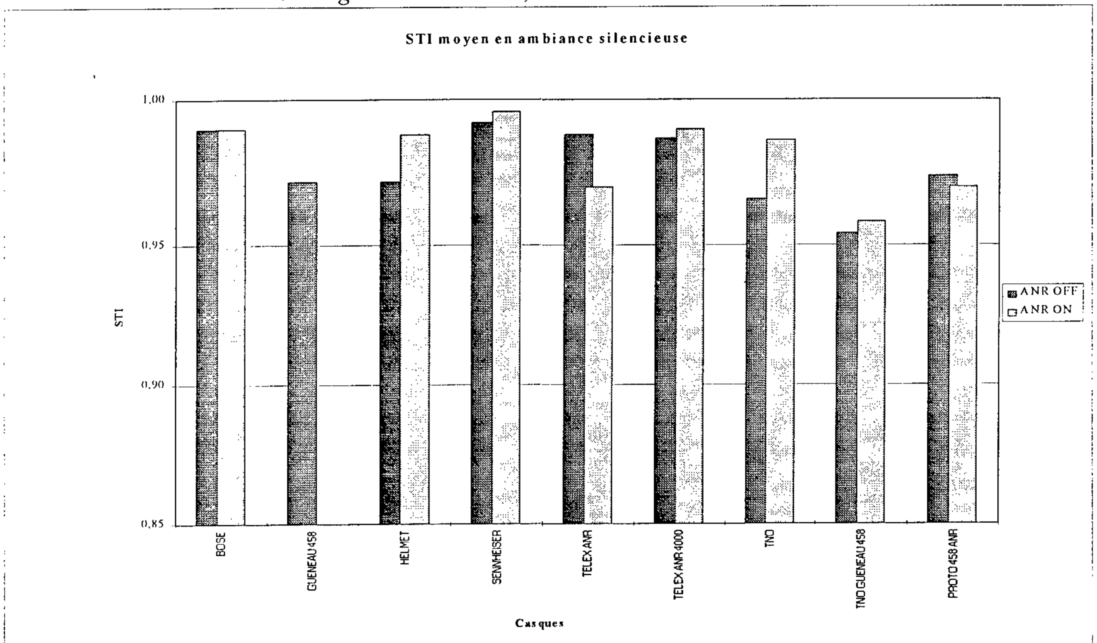
IMASSA 2

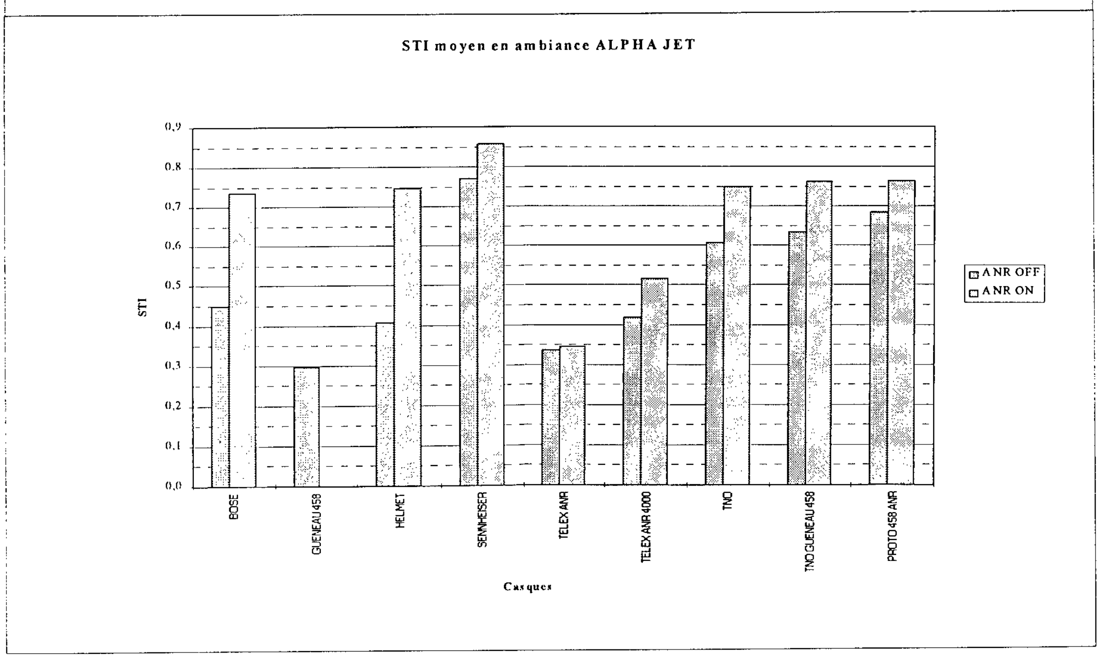
3.3 RESULTATS INTELLIGIBILITE OBJECTIVE

Les expériences ont été réalisées avec un volume sonore de 83 dBA (± 1 dB, compte tenu du réglage du volume par pas de 1,5 dB) pour le signal de test, et ce bien évidemment dans les deux modes de fonctionnement (passif seul, passif + actif) et pour chaque sujet.

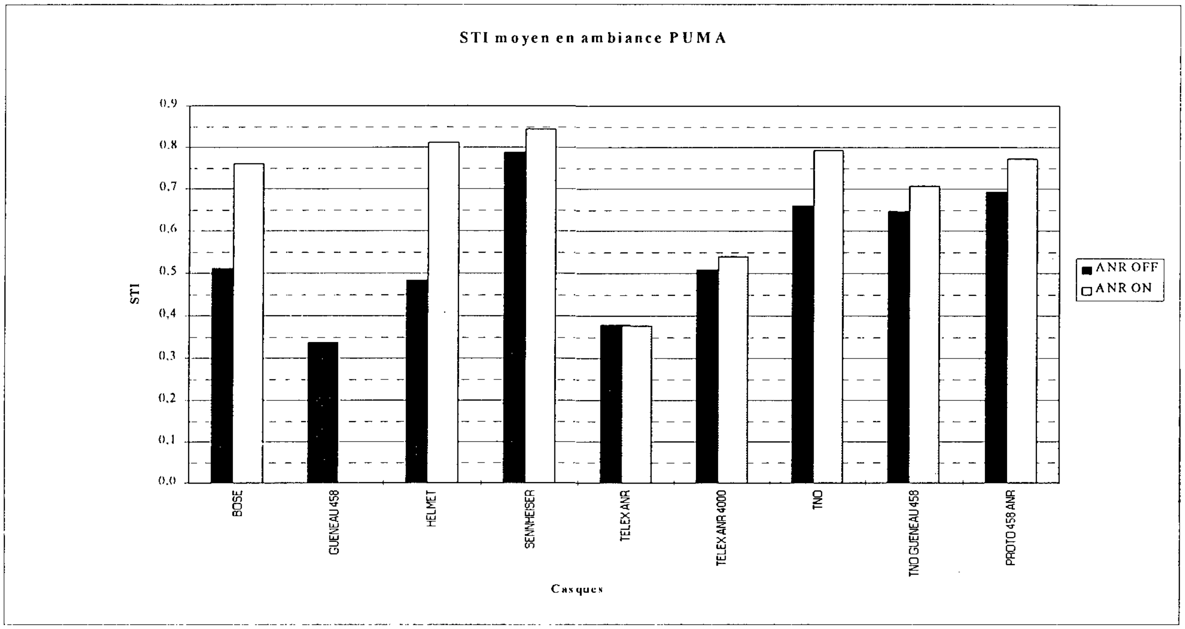
MOYENNE STI	SILENCE		ALPHA JET		PUMA		ROSE	
CASQUES	ANR OFF	ANR ON	ANR OFF	ANR ON	ANR OFF	ANR ON	ANR OFF	ANR ON
BOSE	0,99	0,99	0,45	0,74	0,51	0,76	0,31	0,54
GUENEAU 458	0,97		0,30		0,34		0,16	
HELMET	0,97	0,99	0,41	0,75	0,48	0,81	0,29	0,60
SENNHEISER	0,99	1,00	0,77	0,86	0,79	0,84	0,61	0,68
TELEX ANR	0,99	0,97	0,34	0,35	0,38	0,38	0,20	0,22
TELEX ANR 4000	0,99	0,99	0,42	0,52	0,51	0,54	0,31	0,35
TNO	0,97	0,99	0,61	0,75	0,66	0,79	0,45	0,63
TNO GUENEAU 458	0,95	0,96	0,63	0,76	0,65	0,71	0,51	0,61
PROTO 458 ANR	0,97	0,97	0,69	0,76	0,69	0,77	0,54	0,60

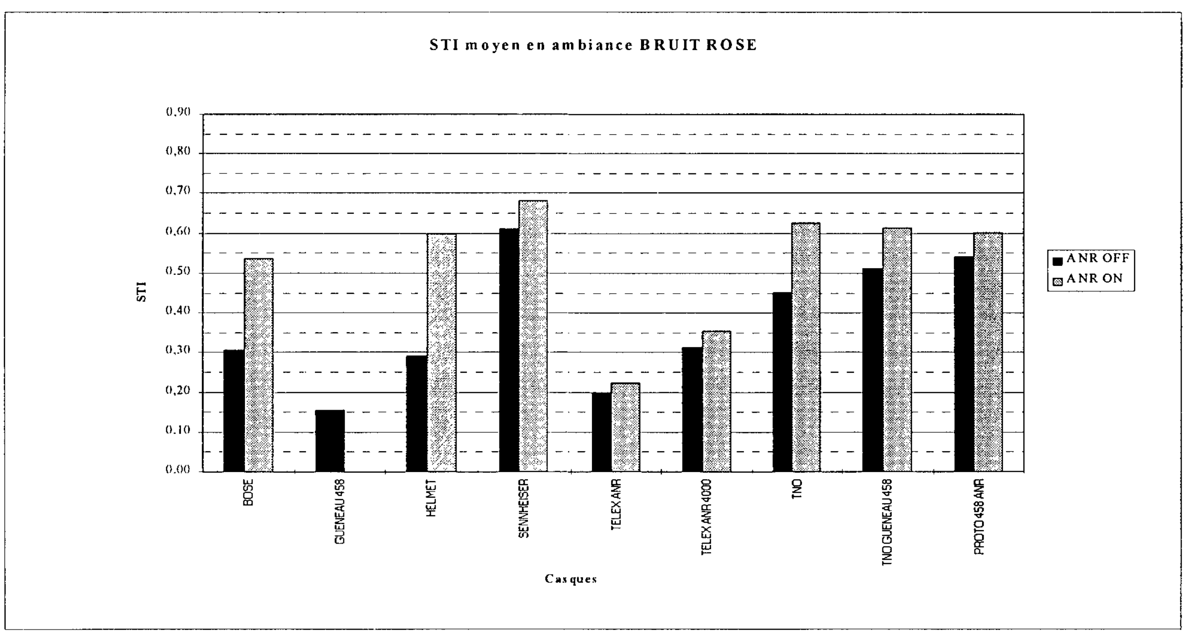
barre gauche: ANR OFF, *barre droite: ANR ON*





barre gauche: ANR OFF, *barre droite: ANR ON*





3.4 INCERTITUDES DE MESURE

Une étude concernant les incertitudes pour les mesures d'atténuations et d'intelligibilité a été effectuée. Pour la mesure d'intelligibilité objective STI sont prises en compte les influences : du niveau sonore à ±1 dB, de l'amplitude du signal retour analysé, du positionnement du casque, de l'impédance de sortie du générateur. De même, pour les mesures d'atténuations, ont été réalisées des mesures de reproductibilité par sujet : de positionnement du bouchon microphone et du casque, de positionnement du sujet. Pour le STI, les scores individuels sont affirmés valides à ±0,06 soit 6%. Pour les atténuations individuelles, les valeurs globales en dBA sont assurées valides à ±1dB. Pour les valeurs en dB par tiers d'octave, elles sont valides à ±2,5 dB.

4.DISCUSSION

ESSAIS EN SITUATION LIMITE :

L'échantillonnage des casques permet d'être confronté à divers comportements en limite de fonctionnement :

- en présence de niveaux très élevés, certains dispositifs sans doute équipés d'une limitation en courant sur les signaux microphoniques peuvent cesser de fonctionner, ce qui signifie que l'influence de la réduction active de bruit est nulle. Ce choix de conception a le mérite de prévenir, certes de façon radicale, tout risque de comportement instable au-delà d'un certain niveau sonore,

- par contre d'autre système continuent de fonctionner mais révèlent des instabilités, dans le sens où le niveau des basses fréquences n'arrive pas à se stabiliser au cours du temps : « le système a du mal à suivre »,

- d'autres systèmes, encore, regénèrent un bruit important, ce qui entraîne l'inefficacité globale de la réduction active de bruit. Il s'agit d'une erreur de conception préjudiciable à la sécurité de l'équipement,

- enfin, des dispositifs fonctionnent correctement sans instabilité, avec une faible regénération de bruit. L'efficacité de la réduction active de bruit est cependant variable d'un système à l'autre en terme de bande de fréquences concernée et de niveau d'atténuation.

Dans l'ensemble, les systèmes répondent correctement en basse fréquence, sans trop de distorsion jusqu'à des niveaux de 105 à 110 dB, ce qui est suffisant pour garantir un fonctionnement correct dans cette zone de fréquences, même dans le cas de bruits cabine présentant beaucoup d'énergie aux très basses fréquences (100 dB environ pour le PUMA autour de 40 Hz).

REDUCTION ACTIVE DE BRUIT ET PROTECTION AUDITIVE :

- l'efficacité de la réduction active de bruit se manifeste de manière significative dans une zone de fréquences couvrant la décade 50-500 Hz. Dans cette zone, l'atténuation active maximale que l'on peut espérer obtenir est de 20 dB environ. Des valeurs comprises entre 15 et 20 dB sont plus courantes sur des plages fréquentielles de largeur 200 Hz

- les phénomènes de régénération de bruit se manifestent généralement entre 1 et 2 kHz et sont limités sur les meilleurs systèmes à des amplifications maximales de l'ordre de quelques dB par bande de fréquences

- en moyenne, les valeurs d'atténuation active dépendent peu de l'ambiance sonore. De façon générale, un casque présentant une bonne complémentarité des protections passive et active peut offrir une atténuation totale de 25 à 30 dB jusqu'à 1 kHz, de 35 à 50 dB entre 2 kHz et 10 kHz

- par rapport à chaque type d'ambiance sonore, il peut être intéressant de qualifier un protecteur auditif par la donnée de l'atténuation globale qu'il peut apporter sur un niveau de bruit pondéré A mesuré au niveau du conduit auditif d'un sujet tête nue. Pour un protecteur donné, l'atténuation passive globale ainsi calculée dépend du type de bruit (essentiellement de sa densité spectrale de puissance) : par exemple, pour un bruit ALPHA JET ou ROSE, elle est environ de 10 dB supérieure par rapport à un bruit PUMA. L'atténuation active globale (différence entre les niveaux pondérés A en mode passif et actif) dépend elle aussi du type de bruit : elle est limitée à quelques dB, en ambiance ALPHA JET, et BRUIT ROSE, elle peut dépasser 10 dB en bruit PUMA. Ainsi, dans de telles ambiances (102 dBA sur microphone de référence), des niveaux de 70-72 dBA sous le protecteur actif peuvent être atteints

- malgré un comportement instable potentiellement dangereux, dans le cas de fuites acoustiques ou de surcharge, certains systèmes fournissent de bonnes performances en utilisation nominale.

INTELLIGIBILITE :

- La correlation observée entre résultats objectifs (STI) et subjectifs (CVC) valide réciproquement la mise en oeuvre des deux méthodes (pour les conditions testées).

- En ambiance silencieuse, l'intelligibilité étant maximale, le STI doit être très proche de 1. Les mesures confirment que tous les casques, en mode passif, présentent un STI supérieur à 0,95. Pour la plupart des casques, le passage en mode actif apporte toutefois une très légère augmentation du STI. Ceci s'explique par le fait que la boucle de réaction vient compenser les

imperfections du haut-parleur. Pour deux casques, au contraire, le passage en actif entraîne une dégradation.

- la corrélation entre intelligibilité et efficacité de la protection passive est vérifiée. En effet, une excellente protection passive permet une intelligibilité excellente, et une protection passive présentant des déficiences ne concède qu'une intelligibilité médiocre ou passable.

- en revanche, l'efficacité de la réduction active de bruit est non nécessairement corrélée à une augmentation de l'intelligibilité, car une fois le dispositif actif en service, l'intelligibilité dépend de la qualité de l'insertion de la phonie et des composants du filtre électronique.

PERSPECTIVES

Un casque à réduction active de bruit bien conçu permet une protection auditive de grande qualité par la complémentarité des protections active et passive aux basses fréquences ainsi qu'un gain d'intelligibilité remarquable.

Un casque prototype fonctionnel équipé d'écouteurs à réduction active de bruit permet d'atteindre une atténuation globale de 20 dB supérieure à celle du casque pilote actuel et une remontée spectaculaire de l'intelligibilité de 35% à 80% sur mots CVC. Ces résultats offrent des perspectives encourageantes d'intégration de la réduction active de bruit ainsi que des critères de conception dans les équipements de tête futurs.

Pour l'utilisateur, l'apport de l'ANR se traduit par une fatigue auditive diminuée à durée d'exposition constante, la possibilité de travailler dans des ambiances plus bruyantes à degré de protection équivalent, une sensation immédiate de confort auditif par la diminution importante des bourdonnements aux basses fréquences et une amélioration importante de la compréhension des messages vocaux. En contre partie, l'utilisation de tels systèmes nécessitera un nouvel apprentissage de l'écoute de l'avion.

REMERCIEMENTS

Les travaux présentés ont été réalisés dans le cadre d'une étude financée par la DCAE / STTE / AVI (Direction des Constructions Aéronautiques, Service Technique des Télécommunications et des équipements aéronautiques).

BIBLIOGRAPHIE

Test of phonetic differentiation : The Rhyme Test, - G. FAIRBANKS, *Acoust. Soc. Am. 30, 596-600, 1958.*

A Multi Language Evaluation of the RASTI-Method for Estimating Speech Intelligibility in Auditoria.- T. HOUTGAST, H.J.M. STEENEKEN, *ACUSTICA 54,4 pp185,199, 1984*

On measuring and predicting speech intelligibility - H.J.M. STEENEKEN, *thèse de doctorat - 1992.*

Objective assessment of speech communication system, H.J.M. STEENEKEN, *Eurospeech Berlin 1993, ISSN-1018-4074-vol1, 1993*

Assessment of active noise reduction systems for hearing protection and speech intelligibility improvement - H.J.M. STEENEKEN,*Rapport TNO-TM, 1994 C-45, 1994*

Description of the STI-measuring method, H.J.M. STEENEKEN, J.A. VERHAVE, T. HOUTGAST, *Rapport TNO-TM, 1994 I-2, 1994*

Diagnostic acceptability measure for speech communication systems, - W.D. VOIERS, *Proc. IEEE ICASSP, Hartford CT, 204-207, 1977*

Effects of Active Noise Reduction on Noise Levels at the Tympanic Membrane

Wagstaff, AS and Woxen, OJ

Royal Norwegian Air Force Institute of Aviation Medicine
P.O. Box 14 Blindern, 0313 Oslo, Norway
Tel. +47 22602790
Fax: +47 22692037

Summary

Active noise reduction (ANR) is an electronic system that works by continuous sampling of noise inside the earshell of the headset with a small microphone. This signal is inverted in phase through the headset speaker, thus reducing noise levels by destructive interference of the acoustic field. The system provides good low-frequency noise attenuation, but air crew differ in their subjective opinion of ANR. The present study is an attempt to provide an objective assessment of the effect of ANR on noise levels at the tympanic membrane. Seven subjects with normal ears were placed in an environment of recorded noise from a BO-105 helicopter. A microphone probe was inserted to within 5 mm of the tympanic membrane of each subjects right ear. Noise levels in the ear were measured without a headset and with two different ANR headsets. Measurements were performed with and without the ANR system on, and, with and without white noise through the headset communication system. The white noise was used to simulate aircraft communication noise.

The two headsets tested had differing levels of passive and active attenuation. The ANR system produced a substantial low-frequency attenuation. However, noise levels in the mid frequencies increased somewhat when the ANR system was switched on. This effect was augmented when white noise in the communications system was introduced, particularly for one of the two headsets. Low-frequency noise attenuation of ANR systems is substantial, but an increased mid-and high frequency noise level caused by the ANR may affect both communication and overall noise levels. Our data provide advice on what factors should be taken into account when ANR is evaluated for use in an aviation operational environment.

Introduction

Noise is an environmental factor in all aviation. It affects flight safety as well as health, exemplified by several airline accidents and incidents caused by poor speech communications. Garbled voice transmission was recently blamed for the much publicised shootdown of USAF Capt. Scott O'Grady over Bosnia on 2 June 1995 (1). This shows that this is an ongoing problem also in military flight operations. Noise-induced hearing loss is also a well-known concern of aviators. Noise attenuation technology is therefore an important aspect of environmental protection in most aviation environments.

The principle for Active Noise reduction (ANR) was first patented by the German scientist Paul Leug in 1936. However, ANR systems have only become commercially available on a larger scale in recent years. Most large-scale aviation headset manufacturers now include ANR headsets in their inventory. There are numerous potential aviation applications for ANR systems, both civilian and military.

An Active Noise Reduction headset or helmet works by continuously sampling the noise inside the earshell using a miniature microphone. The sampled noise is then inverted 180 degrees by an electronic circuit and reintroduced through the earphone speaker. The destructive interference thus induced, cancels out original noise. However, this cancellation is not perfect. ANR systems are effective only in the low-frequency range.

Considering the potential importance of ANR, few publications exist in the open literature which evaluate such systems in relation to human audition.

Some recent studies have however, evaluated speech intelligibility (2-7) and noise levels including the possible prevention of noise-induced hearing loss (2,6,4,8) , as well as the dampening of leakages from spectacles under the earcup (3). So far, published work on ANR effectiveness has documented rather large effects of ANR systems on overall noise reduction. The level of noise reduction depends, however, on the

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

frequency content of the noise. Documentation of ANR effects on speech intelligibility vary in their conclusions, and different methods have been used. Some studies suggest a clear improvement (4,5,7), others only a slight improvement or no improvement at all (3,6). Effects on speech intelligibility also depend on the frequency content of the environmental noise (2).

There exists no international standard for the measurement of ANR headset systems. ANR systems are, of course, fundementally different from passive-only attenuating systems in that they introduce an extra noise signal. The present project was undertaken in order to determine to what extent ANR technology affects noise levels at the tympanic membrane. Measurements at the tympanic membrane include the substantial resonance phenomena in the ear canal/headset system (9,10). Such effects might be important since resonance phenomena in the outer ear might interact with effects of ANR-induced "anti-noise". Moreover, measurements at the tympanic membrane might be helpful in describing the relative importance of different noise frequencies when using ANR systems for noise attenuation and communication in aviation.

Subjects and Methods

Subjects

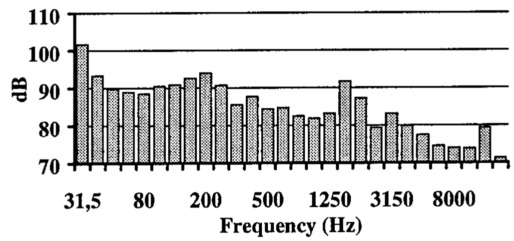
Approval from the regional committee for medical research ethics was obtained.

Eight volunteer subjects were used, and measurements were made on right ears only. None of the subjects had any history of chronic conditions affecting the ear, or present ear-related symptoms or diseases. Clinical otological examination was performed before each test, and was normal for all subjects.

Noise environment

Recorded helicopter noise from a BO-105 helicopter was used inside a sound-proof chamber. The BO-105 cockpit noise has a large part of the noise energy in the low-frequency region, and should therefore be a noise environment suitable for ANR use. In addition, BO-105 turbine noise extends into higher frequencies. Fig. 1 shows the relative frequency content, exhibiting the narrow high-intensity bands typical of helicopter noise. Overall noise levels were monitored throughout each individual experiment.

Figure 1. BO-105 Cockpit Noise.
1/3 octave spectrum.



Measurements and Methodological Considerations

The measurements of noise levels inside the ear canal were performed using a Rastronics PortaRem 20 "Insertion Gain" analyser (Rastronics, Mejeribakken 10, Femstykket 6, DK-3540 Lynge, Denmark). This is a device originally designed for evaluation and fitting of hearing aids. A validation of this measurement system for occupational noise attenuation has been published by Woxen and Borchgrevinck (12), showing this method to be well suited for comparing levels with and without attenuation. However, such levels cannot easily be extrapolated or compared to commonly used levels of reference, for instance a particular dBA level.

A thin silicone tube coupled to a miniature microphone was placed by an ear, nose and throat specialist near the tympanic membrane. First, the tube was inserted into the ear canal until it lightly touched the tympanic membrane. Subsequently, the tube was pulled back slightly, to a position within 5 mm of the membrane to avoid irritation. The tube was then secured to the pinna and placed inferiorly between the tragus and the earlobe, thus causing a smallest possible leakage under the headset seal. The remainder of the tube was secured on the side of the neck. The subjects appeared reasonably comfortable once the tube had been inserted. Substantial care was taken in order to avoid movement of any part of the tube during the experiment.

Two different headsets on loan from the manufacturers were used for this experiment.

Nine measurements in all were performed on each subject. All measurements were performed in the environmental noise previously described. In addition, white noise from a Madsen Midimate 330 audiometer (Madsen electronics, 20, Vesterlundsvej, DK2730, Herlev, Denmark) was introduced into the intercom system during the last 2 measurements, aiming at understanding how noise is affected by the ANR systems. In each headset, the white noise was adjusted with the ANR system turned off by the experimenters to subjectively mimic moderate radio noise. The same white noise levels were used with all subjects.

The experiment was conducted in the following way for each subject:

1. Without hearing protection
2. Headset 1 - passive attenuation only (i.e.ANR system turned off)
3. Headset 1 - ANR system turned on
4. Headset 1 - passive attenuation with intercom noise added
5. Headset 1 ANR system turned on with intercom noise added.

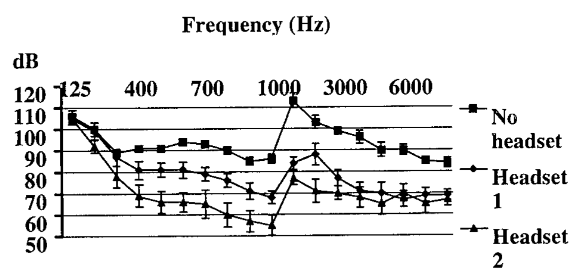
Points 2-5 were then repeated with headset 2.
Each headset was not moved or adjusted on the subjects head between measurements.
Results are in the form of "sweep" levels, i.e. continuous noise levels in decibels for different frequencies. Levels were subsequently entered manually into a computer program for analysis.

Results

The median ambient noise level in the laboratory was 99,8 dBA with a standard deviation of 0,77 dBA.

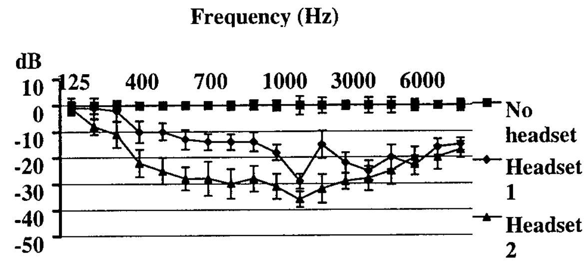
All measurements show clear resonance phenomena peaking at frequencies around 2-3 kHz. Note the height of the peak at these frequencies indicating the relative importance of noise levels in this frequency region. This is somewhat lower than the resonance frequency of the ear commonly quoted around 3000-4000 Hz. However, this can be explained by the presence of the headsets in most of the measurements and the positioning of the probe microphone. Taking these factors into account, our results correspond well to recent advanced technique transfer function measurements performed by other workers (9,10).
Noise levels at the tympanic membrane without and with the two tested headsets are shown in Figure 2.

Figure 2. Noise levels measured at the tympanic membrane for different frequencies - headset 1 and 2 - ANR system switched off. Median values with 95%confidence intervals are shown.



The same measurements, but now showing the no-headset situation as 0 to obtain relative, or actual attenuation levels, are shown in Figure 3.

Figure 3. Attenuation of headsets 1 and 2 at the tympanic membrane when noise level without attenuation is corrected to 0 - ANR system switched off. Median values with 95% confidence intervals are shown.



Clearly, quite large differences in passive attenuation between the two headsets are seen. Headset 2 shows better attenuation in most frequencies with a difference in median levels of up to 16 dB.
The measurements for the lowest frequencies probably show a slightly poorer attenuation for both headsets than if the probe microphone were not fitted. However, earlier measurements on noise leakages performed in our laboratory (11), suggest that this leakage is only occurs at frequencies below 500 Hz and would be, on average, less than 5 dB. This should not affect our results in a decisive way.

Active and passive attenuation levels for headset 1 and headset 2 respectively are shown in figures 4 and 5, and a comparison of the two headsets with ANR systems on is shown in figure 6.

Figure 4. Attenuation of headset 1 with and without ANR. Median values with 95% confidence intervals.

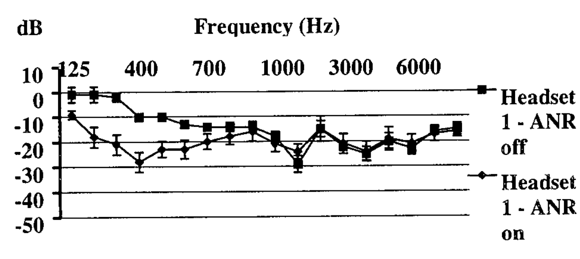


Figure 5. Attenuation if headset 2 with and without ANR. Median values with 95% confidence intervals.

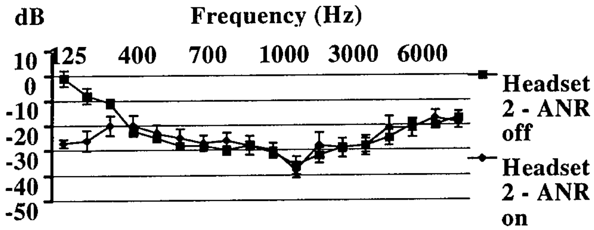
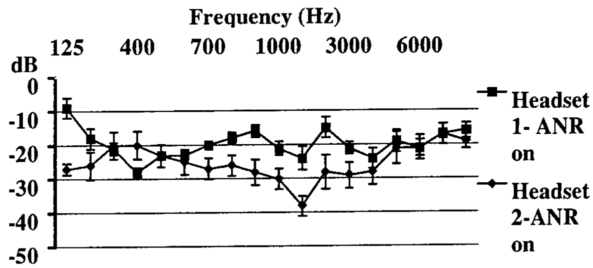


Figure 6. Attenuation of headsets 1 and 2 with ANR switched on. Median values with 95% confidence intervals.



Headset 1 has an ANR system that works over a broader frequency region than than that of headset 2. However, combined passive/active attenuation of headset 2 is still substantially better. Observing different frequencies, it is clear that the ANR systems in both headsets produce a substantial additional attenuation in the low frequency region.

However, in the mid-frequency region, there is a tendency for some amplification of noise. Since this effect had been noted during preliminary studies, the addition of white noise through the intercom should make any such effect more marked.

Measurements made using white noise through the intercom are shown in figures 7 and 8 for the two headsets respectively.

Figure 7. Attenuation of headset 1 with and without ANR, including intercom noise. Median values with 95% confidence intervals.

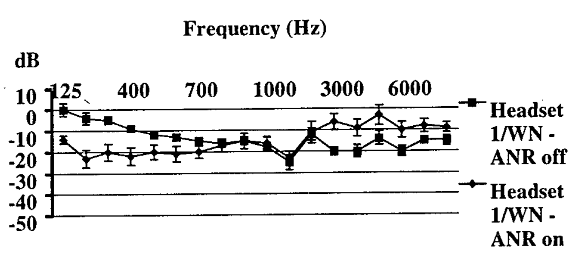
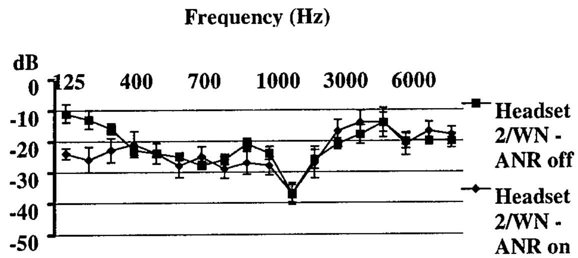


Figure 8. Attenuation of headset 2 with and without ANR, including intercom noise. Median values with 95% confidence intervals.



Here, the low-frequency attenuation of the two ANR systems are similar. However, the introduction of white noise through the intercom has a dramatic effect with ANR systems turned on for headset 1. As one can see from the figure, mid-and high-frequency levels are amplified greatly in this situation. However, any such effect in headset 2 is only a minor one in this experiment.

Discussion

The median ambient noise level in the laboratory was 99,8 dBA with a standard deviation of 0,77 dBA. This corresponds well with the mean cockpit noise level of 100.9 dBA for 65 rotary-wing aircraft published by Gasaway in 1987 (13). The small variability in noise levels rules out any significant adverse effects on results from variations in the noise environment.

We believe that the method employed in this experiment provides a good platform for an analysis of changes in noise levels when using ANR systems.

The variation between subjects is reasonably low (Fig. 2). However, this inter-subject variability of the measurements is lower for the measurements without a headset. The greater variability in the headset measurements is probably due to the fitting of each headset to each person. Thus, a part of the variability of these measurements is probably not due to inter-subject differences alone, but the subject/headset combination for each individual fitting. This corresponds well with earlier work (11), and underlines the importance of not moving the headset between each measurement.

The variability of the measurements is more marked around the peak resonance for the ear/headset system (Fig. 2). This variability may be partly due to differences in ear canal resonance for different subjects. The fact that ear canal resonances vary markedly is well documented by other workers (9) , and could partly explain why different pilots report different subjective benefits of ANR.

The rather large difference in passive attenuation of the two headsets studied (Fig. 3) shows the relative

importance of passive attenuation is no less in ANR headsets. This might indicate that putting electronic circuits into earshells can cause removal of damping materials and thereby decrease passive attenuation.

Active attenuation provides a substantial extra attenuation in the low frequency region for both headsets (Fig. 4 and 5). Headset 1 seems to provide a larger ANR effect over a wider frequency range. However, this does not overcome the better passive attenuation properties of headset 2, shown by the comparison provided in Figure 6. Clearly there are large variations in passive attenuation between ANR headsets. Furthermore, passive attenuation is, in our experiment, more important than active attenuation for the overall attenuation properties. Passive attenuation properties of ANR headsets should therefore be considered carefully when selecting such equipment.

The fact remains that ANR headsets are mainly of use in environments where low-frequency noise predominates. This includes many aircraft environments, such as, for instance, piston-engined aircraft and helicopters. However, some aircraft environments would clearly benefit more from ANR than others.

In addition to the substantial low-frequency damping effect of ANR, there also seems to be some amplification effect on mid-and possibly higher frequencies. This phenomenon which has been observed by other workers as well (2) may be due to a certain summation of waves that have not accurately been phase-inverted. It may also in part be due to the audio system in the ANR headset. The effect is not great, as demonstrated by Fig. 4 and 5, but it may be important, depending on the frequency distribution of the noise. What may be an important point in this context is that there is a large variability in our measurements in the mid-frequency range, in this amplification effect as well. Difference in resonance phenomena between humans (9) and ear-headset systems (10) may mean that a small amplification in certain frequencies in one persons ear might become a larger effect in a different persons ear. This effect might explain differences in subjective opinion of ANR headsets among air crew.

Our measurements of noise levels with additional noise through the intercom (Figs. 7 and 8), show interesting results which also may add to the importance of the above mentioned points. In headset 1, the rather substantial amplification of the intercom noise appears to be due mostly to audio amplification. This follows from the large difference in this amplification effect with and without the white noise. However, an amplification of such a magnitude is not present in headset 2 (Fig. 8). The amplification, or any possible augmentation or change of audio signals through the intercom by ANR systems, must be taken into account when measuring or assessing speech understanding using such headset systems. In previous work, this has not been clearly addressed.

Obviously, the effect described would change the signal-to- noise ratio in a given environment. Hence, subjective opinions of improved speech understanding using ANR headsets may in some cases be due to amplification in speech signals, not overall noise attenuation *per se*.
Again, the importance of this effect may differ substantially from person to person depending on, among other factors, ear canal resonance. In a communication-rich environment, the effect of an ANR effects on communication level and frequency content might be a very important factor to document well before the decision is made whether to introduce ANR into a given operational environment. In this context, the importance of the type of aircraft operations where ANR systems are to be used, should be emphasised. This operational aspect particularly applies where communication may be non-standard and over multiple radio systems, as is in the case in, for instance, search and rescue (SAR) operations.

There are clearly individual differences between people, not only in physical properties such as ear canal resonance, but also hearing levels and personal preference. Current ANR systems can always be turned off and work as ordinary headset/helmet systems. However, maybe a positive future development would be "tuneable" ANR systems, where level and frequency of the ANR effect are adjustable to fit the individual. We do not know at present whether this is a feasible development, but it might increase efficiency of such systems for air crew.

Conclusions

Active noise reduction is a relatively recently implemented technology which, for the first time in noise attenuation, employs the addition of sound to an existing noise field. We have shown that, with ANR, noise field changes in the ear canal may not only include attenuation, but also amplification of sound. The present findings do, in our view, provide data which demonstrate that ANR systems require careful assessment in relation to any operational environment to which implementation is envisaged.
Considering the above, the decision whether to implement ANR in an aircraft operation would depend on four main factors, namely:

1. Level and frequency content of environmental noise

2. Overall attenuation properties of the system, i.e. both active and passive attenuation

3. Intercom effect of ANR system in relation to communication environment

4. Individual (crew factors)

These four points should be taken into account when selecting noise attenuation systems where Active Noise Reduction may be employed. In the longer term, standardised measurement/assessment techniques for ANR systems should be developed.

Acknowledgements

This work was conducted with support from Norwegian Air Ambulance. The Authors wish to thank Harald T. Andersen for help in the planning and evaluation of the experiments.

References

1. Starr, B. Garbled transmission led to F-16 shootdown. Jane`s Defence Weekly 1995 Jul 22:8.

2. McKinley, RL; Steuver, JW; Nixon, CW. Estimated Reductions in Noise-induced Hearing Loss by Application of ANR Headsets. Axelsson. Scientific Basis of Noise-Induced Hearing Loss. New York: Thieme; 1996. pp. 347-60.

3. Rose, DM; Blank, W; Pongratz, H (GAF Institute of Aviation Medicine, Division 4- Ergonomics). Effects of Active Noise Reduction for Cockpit Personnel Wearing GlassesASMA meeting1992.

4. Williams, CE; Maxwell, DW; Thomas, GBSound Attenuation and Speech Intelligibility of a Modified Active Noise Reduction (ANR) Headset for use by P3-C Sensor Operators. Report No.: NAMRL Technical Memorandum 95-2.

5. James S, Harpur K. InFlight Assessment of a Helmet Mounted Active Noise Reduction System in Sea Harrier FRS1. Technical Memorandum MM33, Royal Aircraft Establishment, Farnborough, Great Britain 1990 Jan.

6. Gower DW, Casali JG. Speech Intelligibility and Protective Effectiveness of Selected Active Noise Reduction and Conventional Communications Headsets. Hum Factors 1994;36(2):350-67.

7. Matschke RG. Communication and noise. Speech intelligibility of aircraft pilots with and without electronic compensation for noise. HNO 1994;42(8):499-504.

8. Woxen, O. J.; Borchrevink, H. M. (HQ defence command Norway, FSAN Dept. Environmental Medicine, Oslo.). Acoustic Ear Canal Attenuation of Hearing Protectors and ANR-headsets at >85 dB, Inflight and in the LaboratoryNoise as a Public Health problem. Proceedings of the 6th International CongressJuly 5-9; Nice, France. Arcueil Cedex, France: INRETS; 1993.

9. Møller H, Sørensen M, Hammershøi D, Jensen C. Head-Related Transfer Functions of Human Subjects. Journal of the Audio Engineering Society 1995 May;43(5):300-21.

10. Møller H, Hammershøi D, Jensen C, Sørensen M. Transfer Characteristics of Headphones Measured on Human Ears. Journal of the Audio Engineering Society 1995 Apr;43(4):203-17.

11. Wagstaff A, Tvete O, Ludvigsen B. The Effect of a Headset Leakage on Speech Intelligibility in Helicopter Noise. Aviat Space Environ Med 1996;67:1034-1038.

12. Woxen OJ, Borchgrevink H. Attenuation of Hearing Protectors at 85 dB SPL Investigated by Commercial "Insertion Gain" Method. Scand Audiol Suppl 1991;34:145-55.

13. Gasaway D. Noise Levels in Cockpits of Aircraft during Normal Cruise and Considerations of Auditory Risk . Aviat. Space Environ. Med. 1986;57(Febuary):103-12.

Personal Active Noise Reduction with integrated Speech Communication devices: development and assessment

H.J.M. Steeneken and J.A. Verhave
TNO Human Factors Research Institute
P.O. Box 23, 3769 ZG Soesterberg
The Netherlands

1. SUMMARY

Active noise reduction is a successful addition to passive eardefenders for improvement of the sound attenuation at low frequencies. Assessment methods are discussed, focused on subjective and objective attenuation measurements, stability, and on high noise level applications.

Active noise reduction systems are suitable for integration with an intercom. For this purpose the intelligibility in combination with environmental noise is evaluated.

Development of a system includes the acoustical design, the feedback amplifier, and the speech input facility. An example of such a development is discussed. Finally the performance of some commercial systems and a laboratory prototype are compared.

2. INTRODUCTION

Active noise reduction is an effective tool to increase the sound attenuation of hearing protectors. Especially for the low frequency range the passive sound attenuation of an earmuff or earplug is often insufficient. Active noise reduction can provide an additional attenuation of 20−30 dB at low frequencies (below approximately 500−1000 Hz).

Two specific active noise reduction systems have been developed, one system based on an earmuff and a second system based on an ear plug. The earmuff based system offers a high additional attenuation (up to 25 dB) and can be used with very high noise levels (up to 160 dB SPL).The earplug based system is small and can be used in combination with a gasmask or with a pilot helmet.

A specially developed speech interface allows for injection of speech signals from an intercom system, thus offering a high intelligibility due to the improved sound attenuation and the low acoustic distortion.

Assessment methods of ANR systems differ from methods as used for passive hearing protectors. Due to noise introduced by the electronic system no measurements at the threshold of hearing can be performed. Objective and subjective assessment of the attenuation and the speech quality will be discussed.

3. PRINCIPLE OF ACTIVE NOISE REDUCTION

Active noise reduction is based on the addition of a secondary sound signal to a primary sound signal which has to be suppressed (Lueg, 1936). If the waveform of the two signals are identical but in anti-phase the resulting sound will be zero. A perfect match is theoretical; in practice a feedback loop is used according to the block diagram given in Fig. 1.

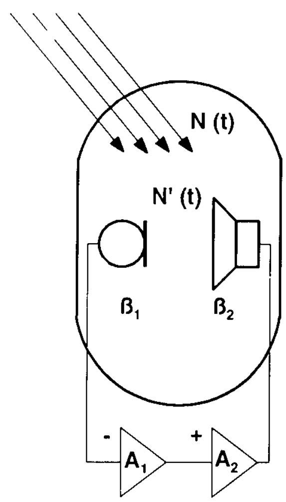


Fig. 1. Schematic diagram of an active noise reduction system within the shell of a hearing protector.

The resulting noise signal N'(t) at the microphone position is the sum of the primary noise signal N(t) (leaking from the outside of the hearing protector) and the secondary compensation signal from the loudspeaker. The latter signal is equal to the resulting noise signal at the microphone multiplied by the loop gain, hence:

$$N'(t) = N(t) - N'(t) \cdot \beta_1 \cdot \beta_2 \cdot A_1 \cdot A_2 \qquad (1)$$

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

$$N'(t) = \frac{N(t)}{1 + ß_1 \cdot ß_2 \cdot A_1 \cdot A_2} \qquad \textbf{(2)}$$

Where ß represents the frequency transfer and the efficiency of the electro acoustic transducers (microphone $ß_1$, and telephone $ß_2$), A_1 represents the gain and frequency transfer of the correction amplifier and A_2 the gain and frequency of the telephone amplifier. The amount of suppression is given by the denominator of equation 2. An increase of the loop-gain ($ß_1 \cdot ß_2 \cdot A_1 \cdot A_2$) results in more suppression.

The frequency transfer of the combination of the electro-acoustic transducers and the cavity under the earmuff is limited. An example of the transfer function of the amplitude and phase of such a system is given in Fig. 2. In view of this transfer function three ranges for the denominator of equation 2 are identified:

(1) the denominator is greater than one which results in a suppression of the primary noise,
(2) the denominator is smaller than one but greater than zero which results in an amplification of the primary noise, and
(3) the denominator becomes zero which results in an unstable system which will oscillate.

The last two possibilities should be avoided by either a lower total loop-gain or correction of the amplitude and phase response. Such a correction can be obtained by a compensation network to be included in amplifier A_1. Reduction of the total loop-gain results in a smaller amount of noise suppression. Therefore, a careful design of the acoustical properties of the cavity within the earmuff, and careful selection of the transducers with an optimal frequency response is required. The frequency and phase response of the compensation network is defined in relation to the frequency response given by $ß_1$ and $ß_2$. A description of the design criteria is given by Olson and May (1953), Carme (1987), and Nelson and Elliott (1993).

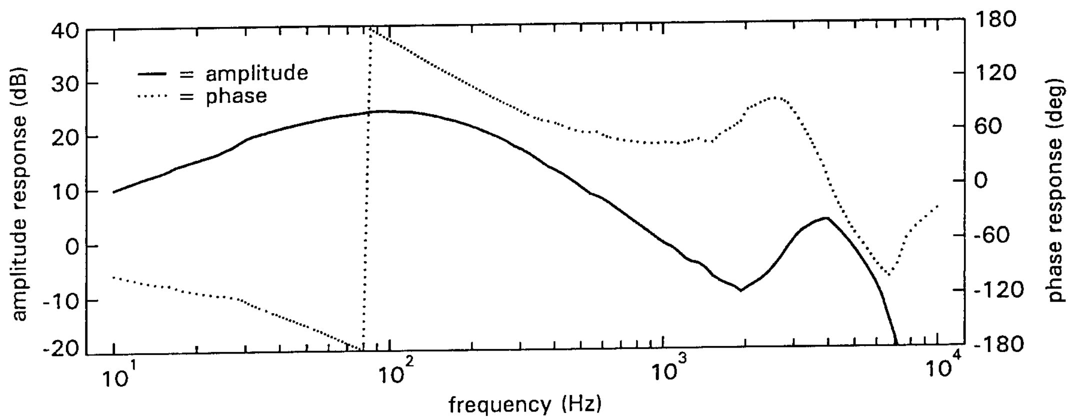


Fig. 2. Amplitude and phase response of the combination of a telephone and microphone within an earmuff placed on the head of a subject.

Insertion of speech signals in the ANR loop is possible. An optimal method to do this is to compensate for the feed back on the speech signal. This can be performed with the circuit given in Fig. 3. The speech signal is then defined by:

$$S(t)' = \frac{A_2 \cdot ß_2 (1+A_1)}{1+ß_1 \cdot ß_2 \cdot A_1 \cdot A_2} \, S(t) \approx \frac{1}{ß_1} \, S(t) \qquad \textbf{(3)}$$

As $ß_1$ (microphone) has a fairly flat frequency response, the speech signal will be reproduced with a nearly flat frequency response.

4. ASSESSMENT OF ANR SYSTEMS

The performance of an ANR system depends on a number of technical properties. Of course the addition of active sound attenuation is a major aspect. However, in order to specify the *personal* protection and safety and not mean values the following items are of interest:

- passive sound attenuation as a function of frequency,
- active sound attenuation as a function of frequency,
- variance among systems,
- variance among users,
- stability on the head,

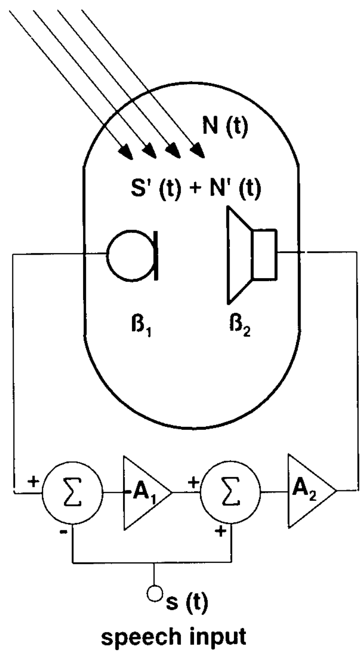


Fig. 3. Schematic diagram of an active noise reduction system including the addition of speech signals.

- stability for open system during placing or removing from the head,
- sensitivity for vibrations,
- maximum sound pressure level (dynamic range),
- overload response,
- speech intelligibility of the integrated communication system.

In this overview we will focus on the sound attenuation and speech intelligibility. However, some examples will be given on the other aspects.

4.1 Sound attenuation

With the introduction of hearing protectors with active noise reduction, which may introduce some system noise at the users ear, the assessment of the sound attenuation according to the standard measuring methods (ISO4869-1) is no longer valid. The ISO method is based on the threshold of perception and, thus, limited to low sound levels. The noise introduced by the ANR systems will interfere with the measurement. Also the sound attenuation of ANR systems may be level dependent. Hence, measurements should be performed at various levels.

Three alternative methods for measuring the sound attenuation are in use:

(1) By comparing the sound pressure level measured under the earmuff with the ANR system switched on and off. The level difference between the two measurement gives the sound attenuation. The measurements are performed by making use of the sense-microphone included in the ANR-loop.

(2) Similar measurements as described under (1) by making use of an additional microphone, positioned close to the entrance of the ear canal.

(3) By subjective matching of the loudness of two sound levels, representative for the additional attenuation of the ANR system.

4.1.1 Objective measurements

The active sound attenuation can be obtained by measuring the difference between the sound pressure levels under the earmuff shell with the ANR system switched on and off. As measuring microphone the loop microphone or an additional microphone placed near the entrance of the ear canal can be used. By means of a positioning system the miniature microphone is placed near the entrance of the ear canal (Fig. 4). This method is called MIRE (Microphone In Real Ear) and is considered to become a new international standard (Technical Committee CEN/TC 159).

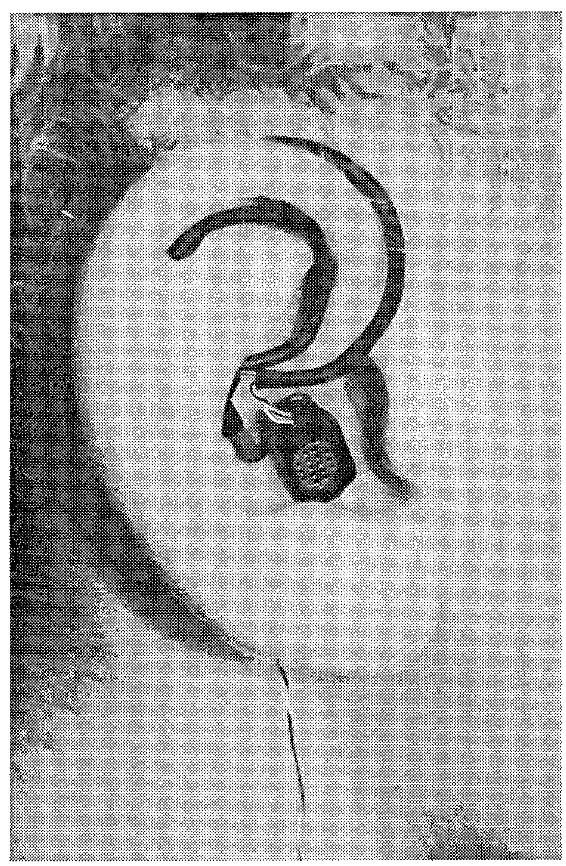


Fig. 4. Measuring microphone near the entrance of the hearing canal.

Preferably, the noise level and spectrum used for the measurements are identical to the noise level and spectrum of the real application. As ANR

systems may have a level dependent attenuation it is advised to determine the attenuation as a function of the noise level.

The attenuation is measured as a function of the frequency. Usually a resolution of 1/3 octave band is used. For this purpose the output signal of the microphone used for the measurements is analysed by a spectrum analyzer.

In order to obtain representative results and to get information on user dependency, various subjects are used.

4.1.2 Subjective measurements

The standardized method for the subjective assessment of the attenuation of hearing protectors is based on a shift of the hearing threshold level if a hearing protector is applied. However, the determination of the hearing threshold in conjunction with an ANR system is not possible as the system itself introduces some noise with a level above the hearing threshold. Therefore, the following method was developed where the subject (with an ANR system for each ear) is placed in a diffuse sound field which alternates periodically between two levels (typically every second). An example of this level alternation is given in Fig. 5.

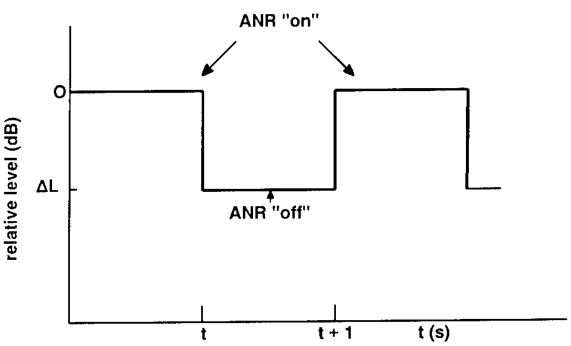


Fig. 5. Relative test signal level as a function of time for the subjective measurements of the suppression of an ANR system. The ANR system is switched on and off simultaneously with the test signal envelope.

During the highest sound pressure level the ANR system is switched on, while during the low sound level the ANR system is switched off. The subject will hear a smaller difference between the two sound levels as the ANR system attenuates only the highest level. The subject is asked to match both levels for equal loudness by adjusting the level difference ΔL between the two signals. The resulting difference in sound level outside the earmuff is equal to the subjective attenuation provided by the ANR system. The adjustment can be made by changing the sound level during the "ANR-off" interval. Since the subject adjusts for a continuous signal, the on/off rhythm is indicated with a light signal. A study (Steeneken and Langhout, 1985) showed that the accuracy lies within $1-3$ dB.

The measurements are performed in a specific room with a diffuse sound field. The test signals that are used consist of noise bands with a bandwidth of 1/3 octave. Measurements are performed in one octave steps. The absolute signal level can be adjusted to any level which is high enough not to interfere with the system noise. However, as the noise reduction of ANR systems may be level dependent, the measurements should be performed systematically as a function of the level.

4.1.3 Comparison of subjective and objective measuring results

A comparison between subjective and objective attenuation measurements was made. The subjective attenuation was measured with four subjects and various signal levels. For one of the conditions the 1/3 octave band signal level was 110 dB SPL. The mean attenuation for these conditions, as a function of frequency with one octave steps, is given in Fig. 6.

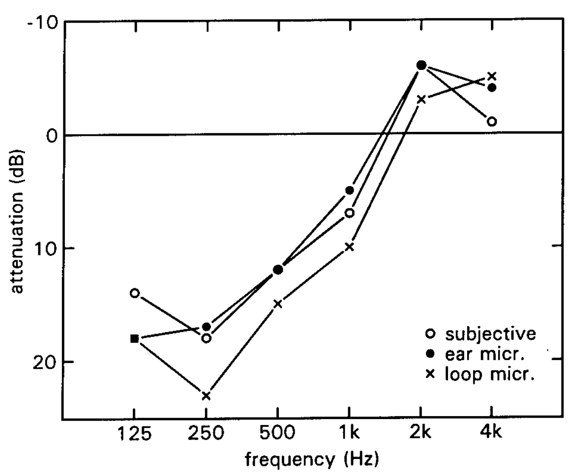


Fig. 6. Mean sound attenuation measured with 4 subjects in one octave intervals for the subjective and objective methods.

The objective attenuation was measured with the loop microphone as well as with a special electret microphone positioned close to the entrance of the ear canal. For the objective measurement a pink noise (level 105 dB SPL) was used.

The results indicate that the attenuation values obtained with the subjective method and those obtained with the ear microphone (MIRE) are in close agreement. The attenuation values obtained with the loop microphone are somewhat higher

(2−5 dB). Obviously, the sound field under the earmuff is not homogeneous and is minimal at the sensing position of the loop microphone.

4.2 Speech transmission quality

The speech quality depends on the method used for the injection of the speech signal. Some systems make use of the method given in Fig. 2 while others inject the speech signal at the sense microphone input. Some designs make use of a correction amplifier.

As the speech transmission quality is defined by the design of the ANR system, the speech injection method, and the suppression of background noise, it is important to assess the speech intelligibility in a representative condition.

This assessment can be done with subjective measures (by making use of speakers and listeners) or by objective methods (by making use of a measuring device). In this study an objective method (the Speech Transmission Index, STI) is used (Steeneken and Houtgast, 1980; IEC 268-16).

The STI is obtained by applying a specific —speech-like—test signal at the audio input and by analysis of this transmitted test signal through the same measuring microphone as used with the MIRE attenuation measurements.

The STI for a specific communication system with ANR as a function of the noise level is given in Fig. 7. The STI is given for two conditions: ANR switched on and off.

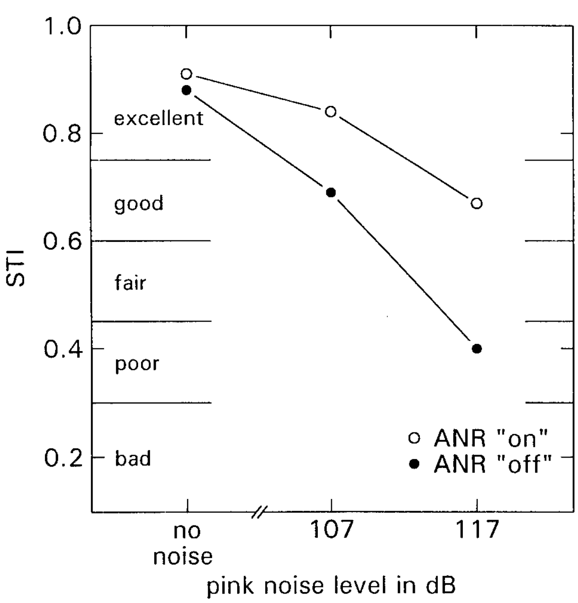


Fig. 7. STI at three noise levels for an ANR system switched on and off. As measuring microphone under the ear-shell the MIRE microphone was used.

Hence, the effect of the ANR on the STI-value can be obtained by comparing the two conditions. Additional to the STI-value also a qualification (based on STI) is given. The improvement of the speech transmission quality is obvious. It is shown that for a constant speech intelligibility (STI=0.7) a 10-dB higher noise level can be applied. Hence the *effective* gain in this situation and for this type of noise is 10 dB.

5. DEVELOPMENT OF AN ACTIVE NOISE REDUCTION SYSTEM INTEGRATED IN AN EARMUFF

The development and assessment of an ANR system can be separated into the following steps:

- acoustical design,
- optimization of the required feedback amplifier,
- development of the speech input facility,
- assessment of sound attenuation and speech intelligibility by objective and subjective methods,
- field trials for conditions with high noise levels e.g. run-up sites of jet aircraft, helicopter and shooting ranges.

5.1 Acoustical design

The acoustical design is of major importance for the final performance of the ANR system.

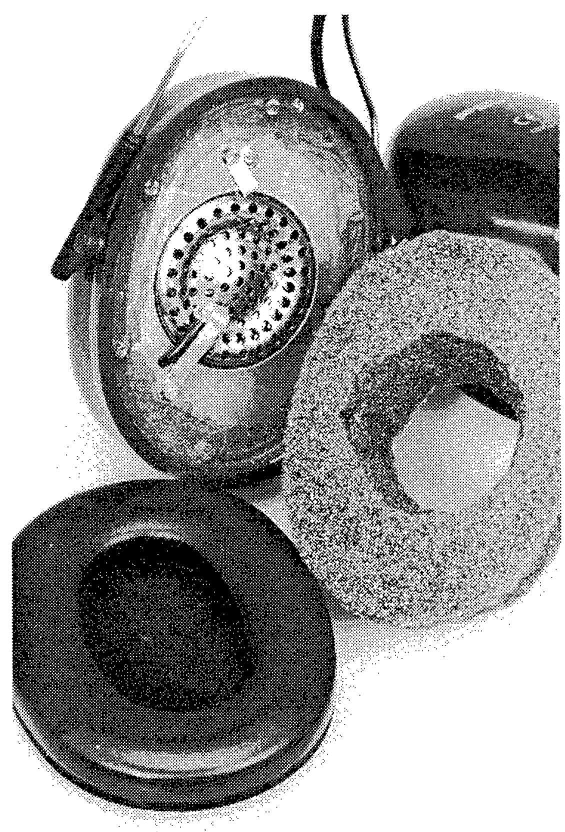


Fig. 8. Lay-out of the acoustical part of an ANR system built in a commercial ear muff.

In order to obtain a high loop-gain within the required stability criteria it is important that a fairly flat frequency response is obtained between loudspeaker and sense-microphone in combination with a smooth phase response with a minimal phase-delay in the required frequency range. In the development stage the acoustical design is considered separately. For a system built into an existing earmuff with commercial transducers (telephone cartridge and electret microphone) the lay-out is given in Fig. 8. The corresponding frequency transfer for the condition that the system is placed on the head of a subject is given in Fig. 9. The figure gives also some indication on individual differences between users as the responses for 14 subjects are given. Analysis of the phase response indicates that a frequency range between 25 Hz and approximately 800 Hz is achievable (phase between ± 60°) with a proper design of the feedback amplifier. The frequency response for the open system (not shown) indicates that the amplitude response drops with 20 dB for the lower frequencies. This guarantees a stable system during the placing on or removing from the head of a user.

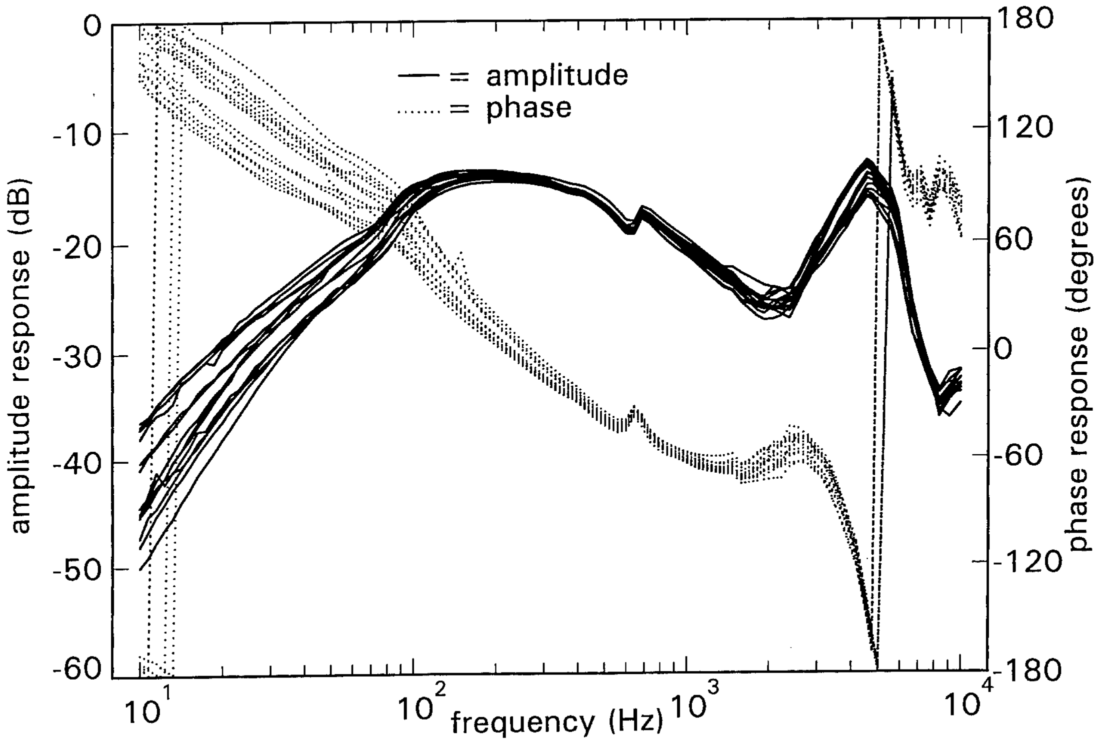


Fig. 9. Frequency response obtained for the electro-acoustical part placed on the head of 14 subjects.

A systematic study on the acoustical design criteria was performed and has lead to a ANR system which can offer an additional attenuation of 25 dB in a specific frequency range.

5.2 Feedback amplifier

The feedback amplifier also provides some filtering in order to define the frequency range in which the total system stability allows a high loop-gain. This can be determined by a Nyquist diagram which gives a vectorial representation of loop-gain and phase. As was discussed in section 3 the nominator of the frequency response should be above zero. This means for the diagram given in Fig. 10 that the area around gain "-1" (0 dB) should be avoided. In the practical situation a stable system is obtained if the indicated area between $-60°$ and $60°$ is avoided.

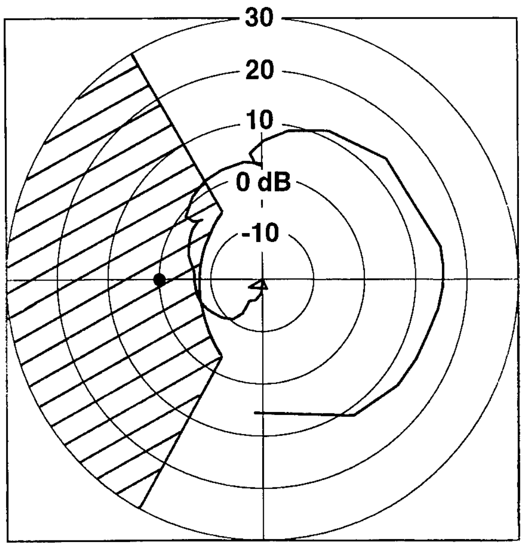


Fig. 10. Nyquist diagram of the total loop of an ANR system.

5.3 System performance

For a typical ANR system, the active sound attenuation is given in Fig. 11. The maximum attenuation is obtained between 50 and 250 Hz, a negative attenuation of 6 dB is obtained at 1000 Hz. This shape is typical for this type of feedback systems.

In order to investigate the individual performance we measured the active sound attenuation for four subjects and for the left and right ear separately. Based on these results the mean attenuation and the corresponding standard deviation were calculated which is also given in Fig. 11. In general the mean attenuation minus on time the standard deviation is used for prediction of the noise dose in combination with a specific noise spectrum. In this example the passive attenuation was not discussed, however, in the same figure the total attenuation (passive and active) is also given.

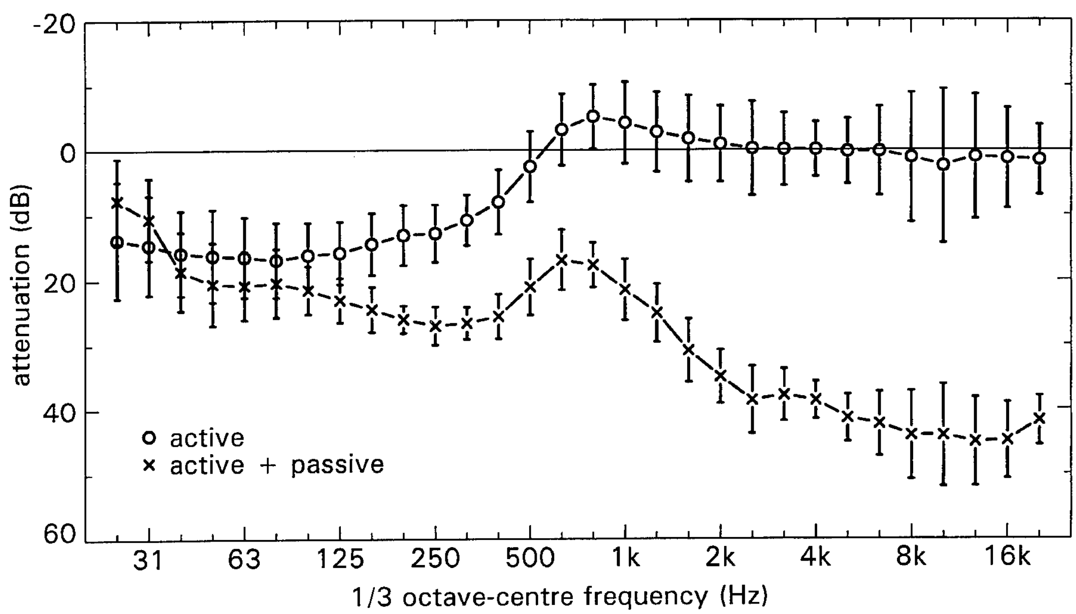


Fig. 11. Mean active attenuation and standard deviation for 3 systems and 4 subjects (8 ears). The total attenuation (passive and active) is also given.

6. DISCUSSION AND CONCLUSION

A comparison of some commercial systems was made. We investigated both the passive and the active attenuation. It was found that the stability of some systems was such that the system started to oscillate when placed on the head of a subject.

Two types of oscillation were found (1) a very low frequent oscillation (below 5 Hz) or above 1000Hz. The sound pressure levels during these instabilities were very high. For this reason we adjust our system 6 dB below the point of instability. If the performance of systems is compared, this security range is often not included. One might get an impression of the stability by observing the amount of negative attenuation. A typical value is 6 dB around 800–1200 Hz. Some systems show a value of over 12 dB. These system are generally not stable. In Fig. 12 a comparison is given for 5 commercial ANR systems (labelled C–G) and the system discussed above (labelled A and B).

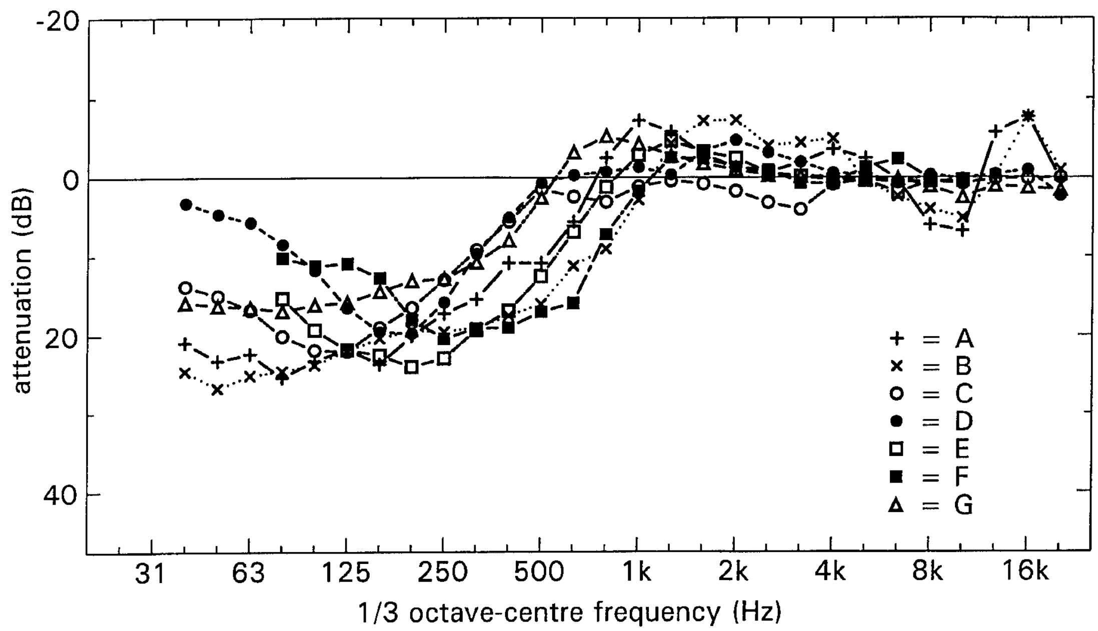


Fig. 12. Comparison of the active and total attenuation of 5 commercial ANR systems. System labelled G is the system discussed in chapter 4.

The curves clearly indicate that most systems provide an additional attenuation of 10−15 dB in a frequency range between 80 and 800 Hz. Only systems A, B and E offer a much higher attenuation. For systems A−B an additional 6 dB stability range is included. This is unknown for the other systems. But the negative attenuation values indicate the same stability.

With respect to the speech intelligibility the performance of the system described above was already given as an example in Fig. 7. Hence an effective gain in signal-to-noise ratio with respect to intelligibility amounts 10 dB. This is determined for a representative noise of a tank.

Recently an active earplug was developed. Such a system is much easier to integrate with an existing helmet (either for a tank or aircraft). The active attenuation amounts 15−18 dB. A study is in progress to improve the performance of this system.

7. REFERENCES

Carme, C. (**1987**). "Reduction active du bruit dans les cavites additives. Méthode et algorithmes de calcul", Journal de Mécanique Théorique et Appliquée, 6, pp.135-160.

Lueg, P. (**1936**). "Process of silencing sound oscillations", US Patent No. 2043416.

Nelson, P.A., and Elliott, S.J. (**1993**). "Active Control of Sound", Academic Press, London.

Olson, H.F., and May, E.G. (**1953**). "Electronic sound absorber", J. Acoust. Soc. Am., 25, pp.1130-1136.

Steeneken, H.J.M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality", J. Acoust. Soc. Am., **67**, 318-326.

Steeneken, H.J.M. (**1983**). "Active noise reduction systems", Report IZF 1983-8 (in Dutch), TNO Institute for Perception, Soesterberg, The Netherlands.

Steeneken, H.J.M., and Langhout, G. (**1985**). "Application of active noise reduction for hearing protection and speech intelligibility improvement", Report IZF 1985-7, TNO Institute for Perception, Soesterberg, The Netherlands.

Steeneken, H.J.M., and Van Velden, J.G. (**1988**). "Application of active noise reduction for hearing protection and speech intelligibility improvement, II", Report IZF 1988-21, TNO Institute for Perception, Soesterberg, The Netherlands.

Adaptive Active Noise Reduction Headset For Helicopter Aircrew

G.J. Pan and A.J. Brammer
National Research Council, Ottawa, Ontario, Canada K1A 0R6
R.B. Crabtree
Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West, P.O. Box 2000
North York, Ontario, Canada M3M 3B9

1. SUMMARY

The feasibility of applying adaptive active noise reduction (ANR) to a communication headset has been explored by applying digital feedforward control to a headset designed for helicopter aircrew. A miniature microphone was mounted on the outside of one circumaural earmuff to provide a reference signal, while the original microphone and earphone located within the volume enclosed by the earcup of a commercial ANR headset were retained to provide an "error" signal and the corrective sound field, respectively. The signals were digitized and processed in real time by a TMS320C31 digital signal processor operating at 40 MHz. The performance of the apparatus has been evaluated in a reverberant room using a recording of Sea King helicopter noise at the aircrew position. The noise was replayed so as to reproduce the sound pressure levels measured in the helicopter during hover. Both noise spectrum and level were confirmed by one-third octave-band analysis. For active control, the helicopter noise was band-limited to from 10 to 1000 Hz. When tested on five subjects, the apparatus controlled the noise at the ear within this frequency range, and the control system was stable. The noise reduction recorded at the error microphone, i.e., close to the ear canal entrance, was in excess of 10 dB from 16 to 300 Hz for all subjects, and ranged from 10 to 26 dB at the rotor blade passage frequency (16 Hz), and from 10 to 20 dB at frequencies up to 200 Hz, depending on the subject. The differences in ANR experienced by the subjects are believed to be associated with variations in the fit of the headset, and remain the subject of continuing research.

2. INTRODUCTION

The high-amplitude, low-frequency noise (~10 to 30 Hz) within helicopters and tracked vehicles remains a challenge to active noise reduction (ANR) technology. In recent years, analog ANR headsets employing feedback control have been developed, and are now commercially available for such environments. An increase in noise reduction of, commonly, at least 10 dB is obtained at frequencies from 40 to 400 Hz (Refs. 1 and 2). However, recent experiments have shown that the analog ANR systems in headsets tend to overload and saturate when operating in high-amplitude, low-frequency noise and infra-sound. This causes the device to generate extraneous noises at the ear, such as clicking and popping (Ref 3). Furthermore, an analog ANR system is usually optimized for a target noise spectrum, so that the gain, phase lag, bandwidth, dynamic range and limiting characteristics of the active controller are set during its design. Although analog ANR controllers are simple to implement, their stability and performance are compromised by their inability to self-adapt their transfer function during operation (Ref. 4), for example, in response to changes in coupling of the sound from the earphone to the ear.

During the last few years, the feasibility of applying adaptive feedforward control to a communication headset has been explored in our laboratory. Although the basic technique is well known and has been successfully employed in industrial applications such as the air conditioning duct, the hearing protector/headset application remains a challenging problem, owing to the small size of the device and the associated requirement for rapid, real-time digital signal processing and control.

There are a number of ways to demonstrate the performance of an ANR system. The easiest is to use numerical simulation. While this may be an attractive research tool, the results may possess little relevance to the performance of working devices. In our previous studies, an acoustic coupler and a KEMAR manikin have been used to test the performance of the system. While the acoustic coupler provides an excellent research tool for software and hardware development (Refs. 5, 6, and 7), the manikin was found to be unsatisfactory for low-frequency ANR tests, owing to its inaccurate representation of sound transmission to the ear canal microphone. For these reasons, human subjects were employed to evaluate headset performance in the present work.

The purpose of this paper is to describe an experimental digital adaptive ANR headset designed for helicopter

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

applications. It should be noted that no communication signals were applied to the headset's earphone during these experiments. In other words, the experimental headset was tested as an ANR hearing protector. The design of the ANR system does, however, permit a communication signal to be passed to the earphone. Also, the experimental device is a one-channel system, and controls the noise in the left earcup only. The control system and algorithms have been developed at the National Research Council of Canada in Ottawa. The experiments were carried out at the Defence and Civil Institute of Environmental Medicine (DCIEM) in Toronto, in order to take advantage of DCIEM's unique noise simulation facility.

3. APPARATUS AND METHODS

3.1 ANR Device

Figure 1 shows the earmuff part of the experimental ANR headset, and Figure 2 shows a block diagram of the digital control system. The headset is of the circumaural type and consists of a passive earcup with cushion, an earphone, and two miniature electret microphones (see Fig. 1).

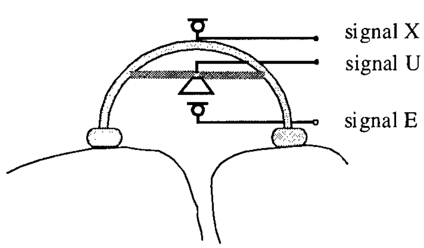


Fig. 1 Earmuff of the ANR headset

The outside microphone, called a "reference" microphone, senses the noise field surrounding the ANR headset and provides a "reference" input signal X to the digital controller (see Fig. 2). The earphone is used as the control actuator, and generates a secondary sound field in the volume enclosed by the earmuff to reduce the noise reaching the ear. The microphone located close to the ear canal entrance is used as the "error" microphone. Its function is to provide feedback to the adaptive controller so that an optimal control signal can be generated to drive the earphone.

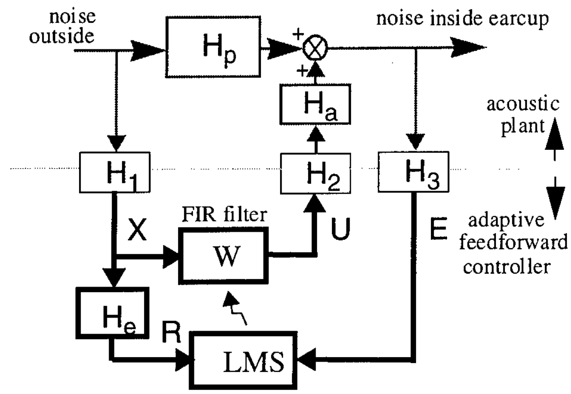


Fig. 2 Control system block diagram

In Figs. 2 and 3, signals X and E represent the outputs of the reference and error microphones, respectively. Signal U is the output of the adaptive feedforward controller, which drives the earphone. The acoustical model of the earmuff from the "reference" to "error" microphone positions is represented by the transfer function H_p. Adaptation of the feedforward controller, W, employs the well-known filtered-X LMS algorithm (Ref. 4). This requires filtering the reference signal to produce signal R by filter H_e, which models the transfer function from the earphone to the error microphone, consisting of H_2, H_a, and H_3. The complete adaptive control system is implemented using a TMS320C31 digital signal processing board equipped with 16-bit analog/digital interfaces.

A multi-rate control structure has been used to provide broad-band, low-frequency noise reduction. This reduces the total signal delay in the control path by increasing the sampling frequency of the A/D and D/A converters and, at the same time, permits the low-frequency performance of the real-time digital FIR filters to be improved. For the measurements reported in this paper, signals were digitized at a sampling frequency of 33 kHz, and the algorithm implementing the adaptive finite impulse response (FIR) digital filter operated at 3.0 kHz. The FIR filter contained 400 coefficients.

3.2 Noise Simulation Chamber

The ANR tests were conducted in DCIEM's large reverberant chamber in which the sound produced by helicopters and tracked vehicles may be reproduced. The rectangular chamber measures 11 x 3 x 6 meters

and meets the sound field requirements of ANSI S12.6 - 1984 (Ref. 8). The helicopter noise was pre-recorded in digital format and reproduced in the chamber by means of a sophisticated sound reproduction system containing band-limiting and parametric filters for spectral shaping. To reduce the possible variation in helicopter noise across experimental conditions, a 2.5-minute segment of the original recording was seamlessly "looped" and re-recorded in order to drive the simulator.

The sound source consisted of multiple electro-dynamic loudspeakers arranged to form a planar array with a radiating area of approximately 7 m^2 (Fig. 3). The array was floor mounted along the shorter wall of the chamber and covered approximately 40% of the wall.

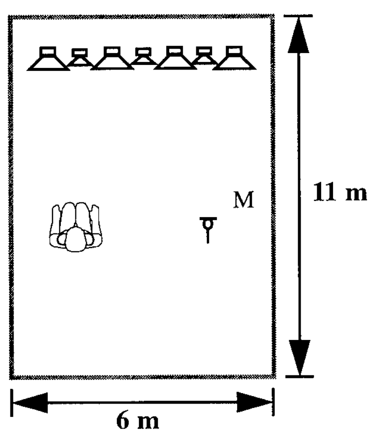


Fig. 3 Sound source and reverberant chamber (plan view)

The geometry of the reverberant chamber results in pairs of positions within the room, equidistant from the array, at which the sound field is matched in sound pressure at frequencies from 10 to 10,000 Hz. For the measurements reported in this paper, the headset was located at one position of such a pair and a microphone (B&K 4149) was located at the other (denoted by M in Fig. 3). The microphone output was monitored in a control room by a B&K FFT analyzer (model 2133) to check both the total sound level and the sound spectrum. This measurement system was totally independent from that employed to measure the performance of our ANR headset (see Section 3.4).

For the active control experiments, the (simulated) helicopter noise was band-pass limited from 10 to 1000 Hz. Although the real source possesses spectral components up to 12 kHz, the noise above 500 Hz is attenuated adequately by the passive earmuff. Thus, the ANR system is required only to attenuate sound at frequencies below 500 Hz. The 10 Hz lower frequency limit of the simulator allows the primary low-frequency helicopter noise source at 16 Hz to be reproduced in the chamber.

3.3 Subjects

Five male subjects, from 21 to 55 years of age, were selected to participate in the experiment. The human subject wore the ANR headset and sat at a preset position in the reverberant chamber at which the simulated helicopter noise was reproduced (Fig. 3). Each subject also wore earplugs. When the test was finished, subjects were asked if they experienced unusual or extraneous noise during the test such as "pops" or "clicks".

3.4 Headset Measurements

Noise Spectra. For the experiments reported here, the noise spectra were measured using a FFT analyzer (Standford Research model SR770). The frequency range of the analyzer was set to be from 0 to 1558 Hz, which results in a frequency "bin" width of 3.906 Hz. For both the reference microphone outside the earcup and the error microphone under the earcup, the spectrum magnitude was the averaged RMS value at each frequency "bin" of the analyzer. In this paper, this magnitude has been converted into the sound pressure level (SPL) in dB (re 2×10^{-5} Pa) using the respective microphone's sensitivity.

For each test involving a human subject, one spectrum at the "reference" microphone and three spectra at the "error" microphone were measured. The reference noise spectrum was usually measured before running the active noise control system, and the level was confirmed by comparison with the B&K analyzer used to monitor the performance of the noise simulator. The three error signal spectra were measured at the error microphone under the following conditions: (1) without active control and without the helicopter noise; (2) without active control and with the helicopter noise and; (3) with active control and with the helicopter noise. The first spectrum provides a measure of the background electronic and acoustic noise. The second spectrum, measured without active control but with the helicopter noise, provides a measure of the passive noise reduction of the earmuff. For circumaural headsets, this spectrum is a low-pass filtered version of the reference noise signal, provided the earmuff is reasonably sealed to the side of the head.

The last and most important noise spectrum is measured with active noise control after the adaptive controller has converged to model the transfer function from the reference to error microphones. This error spectrum provides not only a measure of the active noise reduction, but also any "control spill-over" or noise amplifi-

cation that may be introduced by improper control design, implementation, and/or hardware limitations.

Active Noise Reduction. The active noise reduction was determined by the difference between the noise spectrum measured at the error microphone before the ANR system was switched on and that recorded at the end of a five minute period of active noise control. Note that this is an objective measure of the ANR using a microphone located about 3 cm from the ear canal entrance. The real-ear attenuation of the ANR headset can be determined by using a psychoacoustical method (e.g., Ref. 9).

Note also that the time chosen for running the active noise control system (five minutes) is not determined solely by the rate of adaptation of the control system. The adaptation rate of the filtered-X LMS algorithm is adjustable within a certain range by the "step-size" parameter, and was not optimized in these experiments. While it is desirable to minimize the noise exposure of the subjects to a potentially harmful noise, a sufficiently long time should be chosen for the experiments to ensure that: (1) the adaptive controller has fully converged; (2) the control system remains stable after convergence, and; (3) the ANR reported is measured when conditions (1) and (2) are satisfied.

4. RESULTS AND DISCUSSION

4.1 Noise Spectrum of Sea King Helicopter

Figure 4 shows the Sea King helicopter noise measured at the microphone "M" in the chamber. The solid line shows the 1/3 octave full spectrum of the noise that can be reproduced in the chamber, and the dashed line shows the noise spectrum used for the ANR tests. The latter was obtained by band-limiting the noise from 10 to 1000 Hz.

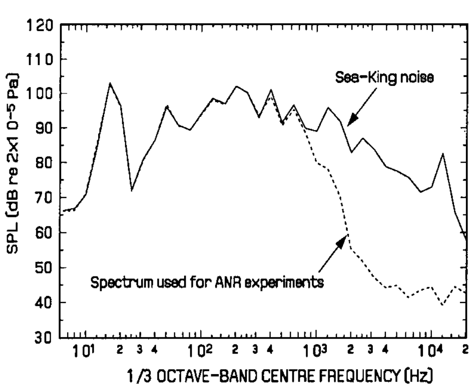


Fig. 4 Spectrum of Sea King helicopter noise

It is clear from Fig. 4 that the helicopter noise has a large amplitude, low-frequency component centered around 16 Hz. This frequency corresponds to the rotation speed of the main rotor blades. Such low frequency noise can be very difficulty to cancel by any ANR system. The difficulties are related to the control filter design and implementation, and the low-frequency capability of the earphone driver.

4.2 Measured ANR on Human Subjects

Figure 5 shows three noise spectra recorded at the error microphone. They were measured: a), with active control and helicopter noise (dashed line); b), without active control and with helicopter noise (solid line), and; c), without active control and helicopter noise (dot-dashed line). For comparison purpose, Fig. 6 shows the same set of measurements performed on a second subject.

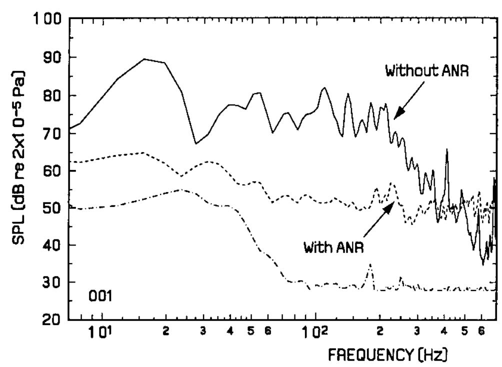


Fig. 5 Noise spectra for subject #1

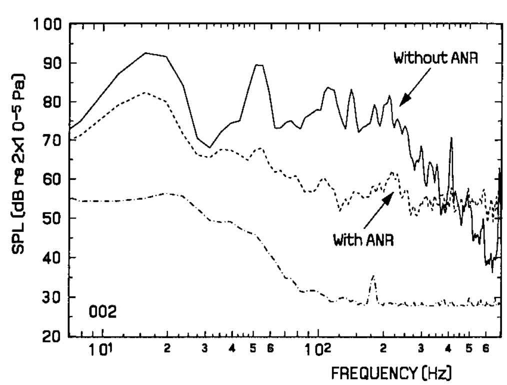


Fig. 6 Noise spectra for subject #2

It can be seen from the solid lines in Figs. 5 and 6 that the noise spectrum under the earmuff without ANR peaks at a frequency of 16 Hz and has a sound pressure level of about 90 dB (re 2×10^{-5} Pa). The noise decreases in magnitude above about 200 Hz with increasing passive attenuation of the earmuff. Thus, the noise that needs be controlled by the ANR system is between 10 and 300 Hz.

The dashed lines in Fig. 5 and 6 show the noise spectrum with the active control system operating. The difference between the dashed and the solid lines shows the ANR performance for this helicopter noise. It is clear from both Figs. 5 and 6 that broad-band active noise reduction has been achieved from about 7 Hz to 400 Hz. The maximum ANR can be seen to occur at about 16 Hz, with a reduction of 26 dB for the subject whose data are in Fig. 5 and 10 dB for those in Fig. 6.

The ANR at other frequencies varied according to the spectrum of the uncontrolled noise, from about 5 dB to 20 dB. Generally, the adaptive controller tends to reduce the noise in such a way that the final residual noise becomes "white" noise. This is demonstrated in Fig. 5 by the relatively flat curve of the noise spectrum.

The mean ANR for the five subjects is shown in Fig. 7 (solid line), together with the spectra for subjects from whom the minimum and maximum ANR was recorded at the rotor passage frequency (dashed-dot and dashed lines). The noise reduction recorded at the error microphone was in excess of 10 dB from 16 to 300 Hz for all subjects, except at a few frequency points where the primary noise level was comparatively low (e.g., 30 Hz). For individual subject, it varied from 10 to 26 dB at the rotor blade passage frequency (16 Hz), and from 10 to 20 dB at frequencies up to 200 Hz.

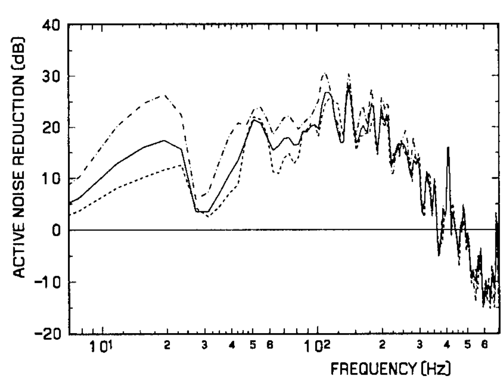


Fig. 7 Average ANR for five subjects

The difference in ANR between subjects can be seen to decrease with increasing frequency. This observation is consistent with incomplete adaptation of the controller when the headset was worn by the subject from whom the least ANR was recorded. This subject reported turning his head from side-to-side during the test, while other subjects tried to remain stationary. Moving the head from side-to-side can be expected to change the seal between the circumaural cushion and the head, or to introduce relative motion between the earmuff and the ear, or both. Either of these phenomena will change the acoustic plant that the adaptive controller is attempting to model. In consequence, the controller, which continually adapts its transfer function, would not have had time to converge prior to the measurement.

4.3 Background Noise

Physiological processes such as the pumping action of the heart, muscular activity and blood flow are associated with the generation of noise in the volume between a circumaural earmuff and the head (Ref. 10). The magnitude of this "physiological" noise increases with decreasing frequency, and has been reported to range from 60 to 70 dB SPL at frequencies below 31.5 Hz (Ref. 10).

The background noise spectrum at the error microphone is shown by the dotted lines in Figs. 5 and 6, and can be seen to be almost frequency independent at 80 Hz and above. At frequencies less than 80 Hz, however, the background noise increases rapidly, reaching sound pressure levels in excess of 50 dB at 30 Hz and below. Such levels are in excess of those attributable to the electronic noise of the apparatus and are believed to be due to physiological noise.

The presence of physiological noise in the volume enclosed by the earmuff will set an upper limit to the ANR at low frequencies. This is because the physiological noise is not correlated with the helicopter noise and is not sensed by the reference microphone in our single-channel, adaptive, feedforward control system.

4.4 Control Stability

It has been observed in other experiments that the noise reduction may oscillate or decay under some conditions of controller performance. That is to say the ANR measured, for example, after the adaptive control is run for thirty seconds may be greater than that measured several minutes later. In extreme cases, the control may eventually become unstable. Such phenomena require the long-time performance of the ANR system to be established, which is an important consideration in our work.

In this experiment, the adaptive feedforward ANR system was found to be stable. Although each subject wearing the ANR headset produced different physiological noise levels and each fitting involved subjective sealing of the cushion to the head, the control system adapted to provide substantial low-frequency ANR for each user.

5. CONCLUSIONS

A digital ANR headset based on adaptive feedforward control has been developed, and the performance measured on human subjects for simulated helicopter noise. The adaptive control system demonstrated good stability and substantial ANR from 12 to 300 Hz. The results confirm the feasibility of applying digital ANR to the helicopter cockpit environment.

A further step towards practical application and commercialization of the technology is to develop a portable prototype of the digital ANR headset, and evaluate its performance in a helicopter. This work is currently under way as part of the continuing collaboration between NRC and DCIEM.

ACKNOWLEDGEMENT

The authors would like to thank Stephen King and Michael Parrell for their help in performing the experiments.

REFERENCES

1. Forshaw, S.E., J.M. Rylands, and R.B. Crabtree, "Evaluating the Effectiveness of Active Noise Reduction in Flight Helmets", DCIEM No. 88-RR-34, August 1988.

2. McKinley, R.L., J.W. Steuver, and C.W. Nixon, "Estimating Reductions in Noise-Induced Hearing Loss by Application of ANR Headsets", in "Scientific Basis of Noise-Induced Hearing Loss", edited by A. Axelsson, et al., New York, Rieme, 1996 (ISBN 0-86577-596-6), pp. 347-360.

3. Crabtree, R.B., "Active Noise Reduction Systems: Their Interaction with Very Low Frequency Acoustical Energy", Proceedings of 12th Triennial Congress of the International Ergonomics Association, 1994 (ISBN 0-9698544-4-7), pp 256-258, Vol. 5.

4. Nelson, P.A. and S. J. Elliot, "Active Control of Sound", London, UK, Academic Press, 1992 (ISBN 0-12-515426-7).

5. Ryan, J.G., E.A.G. Shaw, A.J. Brammer, G. Zhang, "Enclosure for Low-Frequency Attenuation Assessment of Active Noise Reducing Circumaural Headsets and Hearing Protectors", Canadian Acoustics, 21, 4, 1993, pp 19-20.

6. Pan, G.J., A.J. Brammer, R. Goubran, J.G. Ryan, and J. Zera, "Broad-Band Active Noise Reduction in Communication Headsets", Canadian Acoustics, 22, 3, 1994, pp 113-114.

7. Pan, G.J., A.J. Brammer, J. Zera, and R. Goubran, "Application of Adaptive Feed-Forward Active Noise Control to a Circumaural Hearing Protector", Proc. of the 1995 International Symposium on Active Control of Sound and Vibration, Newport Beach, CA., July 1995, pp 1319-1326.

8. "ANSI S12.6 - 1984 American National Standard Method for the Measurement of the Real-Ear Attenuation of Hearing Protectors", New York, American National Standards Institute.

9. Zera, J., A.J. Brammer, and G.J. Pan, "Comparison between subjective and objective measures of active hearing protector and headset attenuation", accepted for publication in J. Acoust. Soc. Am., July 1996.

10. Shaw, E.A.G., and J.E. Piercy, "Audiometry and physiological noise", in "Proceedings of Fourth International Congress in Acoustics", Copenhagen, 1962, Paper H46.

EFFECTS OF ACTIVE NOISE REDUCTION IN ARMOR CREW HEADSETS

B. Wayne Anderson
U.S. Army Research Laboratory
Human Research and Engineering Directorate
ATTN: AMSRL-HR-ML
Fort Monmouth, New Jersey 07703-5630, USA

Georges R. Garinther
U.S. Army Research Laboratory
Human Research and Engineering Directorate
ATTN: AMSRL-HR-SD
Aberdeen Proving Ground, Maryland 21005-5066, USA

1. SUMMARY
The armor environment, like that of aviation, makes communication difficult and often produces a hearing loss in the crewmembers. In an attempt to improve this situation, the Army is presently fielding tankers' helmets with Active Noise Reduction (ANR) as a part of the Vehicular Intercommunications System (VIS). A number of studies were conducted to evaluate the effectiveness of ANR for the armor environment. In-the-ear noise level measures were done and speech intelligibility tests conducted. For armored vehicles producing noise levels of 114 dB(A), these helmets reduce the noise at the ear to 83 dB(A) when the intercommunication system is not keyed, 90 dB(A) when the system is keyed, and 94 dB(A) when the system is keyed with a person talking over the system. This is an improvement in noise reduction of about 17 dB(A) compared to the helmets presently being used. This improved noise attenuation has increased speech intelligibility from 68% to 89%. According to previous studies, such an improvement can be equated to a 25% increase in successfully accomplished armor missions. Incorporation of ANR into these helmets has increased low frequency attenuation by up to 13 dB above the passive attenuation of these helmets. At frequencies greater than 800 Hz, ANR does not provide any additional attenuation above the passive attenuation. The attenuation produced by these new helmets has increased the allowable daily exposure time in armored vehicles from 20 minutes to 12 hours.

2. INTRODUCTION
Although armor is, in many ways, very different from aviation, tracked vehicles and aircraft have two major operational concerns in common—these are poor communications and hearing loss among members of the crew. The problem for both is the level of noise inside the crew compartments. In armored vehicles such as a Bradley Fighting Vehicle or an Abrams Tank, the sound levels range from 98 dB(A) to 117 dB(A) under normal operating conditions with peaks that can exceed 128 dB. Although these levels are a function of crew position, road surface, and speed of the vehicle, they are similar to the noise levels found in military helicopters that can range from 95 dB(A) for the Black Hawk to 115 dB(A) for the Chinook.

The U.S. Army recognized that the intercom system in tanks and other armored vehicles did not provide adequate communications between members of the crew. For this reason, a decision was made to upgrade the intercom system within armored vehicles (Ref. 1). To improve speech communications, this new armored vehicle intercom, the Vehicular Intercommunications System (VIS), included digital circuitry, improved electrical shielding, and voice activation (Ref. 2). At about the same time, Active Noise Reduction (ANR) had been successfully demonstrated, and it appeared to be a viable technology for improving communications while reducing the potential for hearing loss in armor crewmen. The Army, therefore, procured new headsets that included ANR for the new intercom system. Fielding of the VIS began early in 1996.

The VIS was the first large fielding of any military communications system in the world that included headsets with ANR. Because of this, the VIS program office asked the U.S. Army Research Laboratory to collect data for determining the effects of ANR on speech communications and hearing protection. This report is a summary of those data.

3. THE VIS STUDIES
The studies described here were conducted at the Aeromedical Laboratories at Wright–Patterson Air Force Base, Ohio and Aberdeen Proving Ground, Maryland. At Wright–Patterson, data were collected on the ANR systems for attenuation, speech intelligibility, and noise levels at the ear. Additional attenuation data were collected for typical use conditions. These data were obtained with subjects while wearing user equipment such as glasses and a gas mask, and, for the armored vehicle crewmen's helmet, with the chin strap unsnapped. At Aberdeen Proving Ground, impulse noise data were collected during firing of a 155–mm howitzer. Actual steady-state noise levels at the ear were obtained while riding in a Bradley Fighting Vehicle (BFV).

The ANR Headsets Tested
A variety of passive and active noise reduction headsets was

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

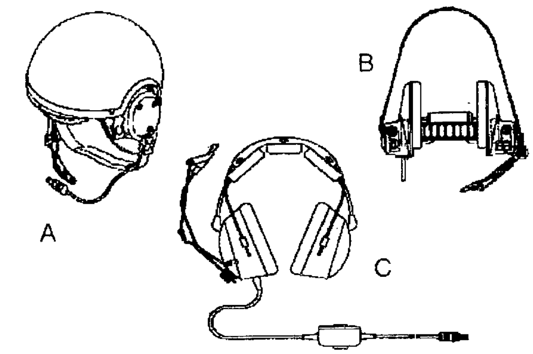


Fig. 1. Headsets tested: (A) Combat Vehicle Crewmen's (CVC) headset mounted in the helmet shell; (B) Communications Aural Protective System (CAPS) and Artillery Communications Aural Protective System (ACAPS) with Velcro strap over the top (front view) for securing to combat helmet; and (C) the Commercial Grade Headset (CGH) with lip-microphone and boom.

provided for the VIS procurement by the prime contractors, Grumman (now Northrop-Grumman) and Royal Ordnance. Of these headsets, the four basic types with ANR were used for the tests (Fig. 1). These were the Combat Vehicle Crewman's (CVC) helmet, the Communications Aural Protective System (CAPS), the Artillery Communications Aural Protective System (ACAPS), and a Commercial Grade Headset (CGH).

The CVC, from Bose Corporation, is the new tanker's headset. It fits into the tanker's helmet shell and includes high attenuation ear seals and an improved noise-canceling microphone. The CAPS and the ACAPS headsets, from Gentex combat infantry helmet and provide both hearing protection and communications capabilities within armored vehicles. They will be used for mounted infantry and command and control vehicles. The CAPS and ACAPS are similar in design except that the ACAPS includes a talk-through circuit that permits soldiers to communicate easily without removing the headset while they are disconnected from the intercom system. The CGH, also from Gentex, consists of ANR earmuffs that can be connected to the VIS. This headset (primarily for command and control) can be used in a noise environment where a helmet is not required.

Attenuation Studies

The attenuation studies were conducted using the microphone–in–real–ear method. Subjects were tested in a reverberant room at a noise level of 115 dB(A). Measurements were made at 1/3–octave bands for three conditions: (1) without the headset, (2) with the headset on the subject, ANR off, and (3) with the headset on the subject, ANR on. Three measurements were taken from both ears of 10 subjects.

For the CVC, the overall average attenuation was 35 dB(A). Fig. 2A shows the average passive and average total attenuation at the octave band frequencies for the CVC. The effect of the ANR for the CVC occurs below 800 Hz which ranged between 2 and 14 dB, the greater attenuation being at the lower frequencies. Between 800 and 6300 Hz, the attenuation was slightly reduced from the passive attenuation by 1 or 2 dB.

The overall average attenuation for both the CAPS and ACAPS was the same, 27 dB, with the 1/3-octave band spectra being within 1 or 2 dB. This result was not unexpected since they are virtually the same design. For this reason, their measurements were averaged and the results are reported together (Fig. 2B). The overall attenuation of the CAPS and ACAPS is less than that of the CVC because of its smaller earcup volume; the design of the CAPS and ACAPS required that the earcups be worn under the infantry helmet. The CAPS and ACAPS show a similar pattern to that of the CVC with most of the attenuation occurring at the low frequencies, although at some of the

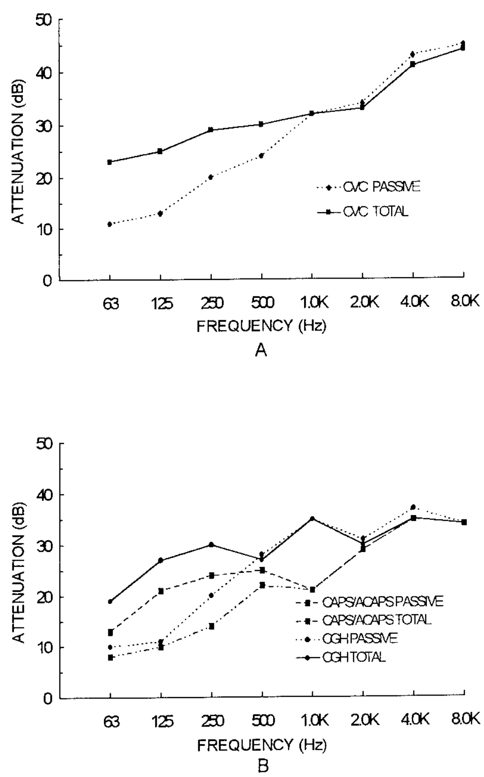


Fig. 2. Average passive and average total 1/3-octave band attenuation levels for the CVC (A); and the CAPS/ACAPS and CGH (B).

higher frequencies, there is a slight increase, 1 to 2 dB, in attenuation over the passive attenuation.

For the CGH, the attenuation curves follow the same pattern as the other headsets with the greatest attenuation at the lower frequencies (Fig. 2B). The attenuation is better for the CGH than for the CAPS and ACAPS, and, in some cases, better than the CVC. Again the reason is believed to be because of the larger earcups which provide greater passive attenuation.

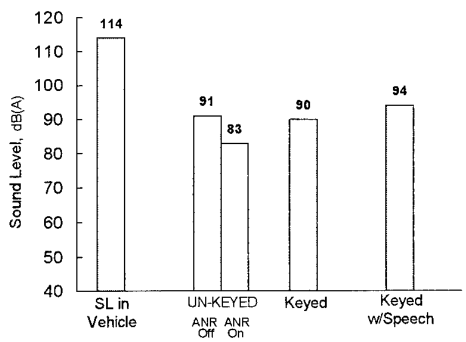


Fig. 3. Comparison of Sound Levels (SL) at the ear for keyed and un-keyed microphone conditions in 114 dB(A) Bradley noise.

Sound Levels at the Ear

Actual noise levels were measured at the ear of a subject wearing a CVC helmet while riding in the Bradley. The vehicle was traveling at 40 mph, that is, about $^2/_3$ of the top speed. A calibrated electret microphone was mounted onto a silicone earplug inserted in the subject's ear. A recording of the noise was made using a DAT tape recorder, and it was analyzed with a $^1/_3$-octave band real–time analyzer.

This test measured noise entering the ear both through the earcups and through the lip-microphone. The results are pictured in Fig. 3. With the overall noise level in the vehicle of 114 dB(A), the level at the ear was 83 dB(A) for ANR on, with the level increasing to 91 dB(A) with ANR off. When the microphone was keyed the level was 90 dB(A) with no one talking, and the level increased to 94 dB(A) when speech was present.

The same test was conducted with 10 subjects in the reverberant room using recorded vehicle noise. The subjects were allowed to set their own listening levels based upon comfort and their ability to understand the speech. As was expected, the sound levels varied somewhat, but the average noise levels were about the same as those found in the earlier study in the Bradley.

Speech Intelligibility Tests

As mentioned earlier, the impetus for the development of the VIS was the poor speech communications in armored vehicles. Speech intelligibility with the old system, the AN/VIC–1, was about 68% as measured by the Modified Rhyme Test (MRT).

A series of speech intelligibility tests for the VIS was conducted, in the laboratory, with noise levels and spectra typical of the environments where the systems would be used. The VIS was set up for a 10 crew-position configuration in one of the reverberant rooms at Wright–Patterson Air Force Base.

Two noise recordings—one from a BFV and the second from the Paladin self-propelled howitzer—were used. The noise levels of the tests were 114 dB(A) for the Bradley and 109 dB(A) for the Paladin. These noise spectra are very high at low frequencies and rapidly fall off at the higher frequencies. Speech intelligibility for all the varieties of headsets was tested in these noise spectra. The results are shown in Fig. 4.

The CVC will be used in the Bradley, the Paladin, and the Abrams tank. Testing was conducted with the Bradley and Paladin noise spectra and intensities. Since the noise level and spectrum for an Abrams tank are similar to that of the Paladin, a separate test was not run for that vehicle. The average speech intelligibility scores for the CVC were 89% for the worst case vehicle, the Bradley. This improved to 92% in the less intense Paladin noise.

Both the CAPS and the ACAPS produced about the same speech intelligibility, 89% and 90% respectively, in the Paladin noise. At the higher level Bradley noise, the CAPS dropped to

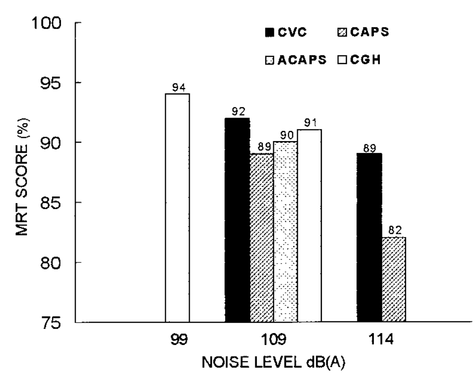


Fig. 4. Speech intelligibility: percent words correct on the Modified Rhyme Test (MRT) for the headsets tested at the different noise levels.

82%. The lower scores of the CAPS and ACAPS compared to the CVC is believed to be a result of smaller earcup volumes of the CAPS and ACAPS.

The CGH will be used in shelters where the noise levels are typically below 100 dB(A). A shelter noise was not available, but with a generator and other equipment running, the noise could be similar in spectrum to that of the Paladin. For these reasons, the Paladin spectrum at 2 levels—99 dB(A) and 109 dB(A)—was used for the test. At the 99 dB(A) level simulating the noise of a shelter, the speech intelligibility score was 94% and, at the 109 dB(A) level simulating the Paladin, the score was 91%.

Effect of ANR on Speech Intelligibility

Since data have been controversial on the effect of ANR on speech intelligibility, an attempt was made to gain some insight into that question. In this test, the problem was confounded by, the desire to simulate vehicle conditions. Subjects were permitted to set their own speech levels to maximize their communications and comfort. Data were then collected for ANR off and ANR on conditions. Since the gain level settings affect the speech level at the ear, and thus the speech intelligibility, subjects were asked to record their settings after each test run. The gain setting was divided into 12 segments, like a clock, so the subjects could mark the number of the segment that matched their setting for that run.

During the early groups of testing, it was found that the subjects frequently adjusted the voice level. The average difference in gain between ANR on and ANR off was 1 segment. Subjects tended to set the level lower when they used the ANR. This amounts to about a 5 dB difference in gain. At the same time, they showed no difference in speech intelligibility.

In later runs of the speech intelligibility tests, the subjects generally stopped adjusting the gain control. On average, there was no difference between gain settings. For these runs, they did show an average difference of 5% improvement in speech intelligibility with ANR on.

Effects on Impulse Noise

Since the VIS headsets were the first ANR headsets to be fielded for rigorous Army use, concerns arose about high impulse noise, that is, above 175 dB, such as a Paladin firing a long range charge. At those levels, there was concern that the ANR might become unstable and create spurious noise that could be hazardous to hearing.

To test this possibility, measurements were made with a manikin head and torso wearing an ACAPS headset, during Paladin firings (Ref. 3). The manikin was seated inside the turret. Impulse noise waveforms from the Paladin were evaluated for ANR off and ANR on conditions.

Results showed that the ANR did not cause spurious noises from the high impulse noise. These data also provided an answer to the question of whether or not the ANR would be able to reduce the peak levels of an impulse noise. It was observed that, in the Paladin, impulse noise levels that were typically above 175 dB produced no reduction in the intensity of the wave as a function of ANR.

Since this study was completed, Dancer (Ref. 4) also looked at impulse noise with and without ANR. He found that ANR does reduce the impulse levels between 100 dB and 150 dB, but the effect becomes smaller as the levels approach 150 dB.

Military Use Studies

Soldiers typically use the headsets in conjunction with other equipment which can affect the attenuation properties. For this reason, a group of studies was conducted to evaluate the effect of typical use on attenuation of the headsets. To collect data on attenuation, the same methodology was used for these studies as for the earlier attenuation tests. The effects of two types of equipment—the gas mask and hood (Mission Oriented Protective Posture, MOPP) and glasses (protective eye-wear)—were compared to the baseline data obtained from the attenuation study.

Fig. 5 demonstrates the findings for the CVC as an example of typical-use effects. As expected, attenuation was degraded from the original baseline data for both of these types of equipment for all the headsets which were tested, that is, the CVC, the CAPS, and the ACAPS. The low frequencies are primarily affected. The temples of the glasses break the seal and allow additional noise to enter the earcups. The gas mask and hood were worse than the glasses since the headset is worn over the hood which fits between the ear and the earcups. This, of course, breaks the ear seal, and results in the loss of much of the passive attenuation. In addition, with the hood located

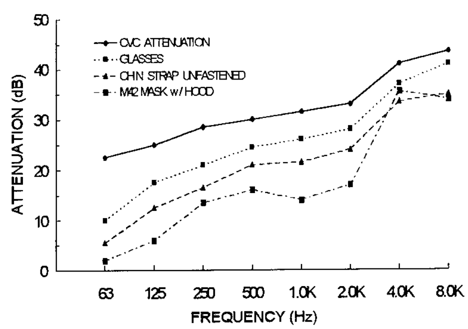


Fig. 5. Average 1/3-octave band attenuation levels for the CVC with various use conditions—baseline CVC attenuation; with glasses; with chin strap unfastened; and with the gas mask and hood.

	CURRENT CVC	VIS CVC	
		TOTAL ATTEN.	PASSIVE ATTEN.
SL at ear (mean atten. minus 1 sd) (dBA)	99.6	83.1	91.3
Allowable Exposure Time (Hrs)	0.3	12.4	1.9
Speech Intelligibility MRT Score (%)	68	89	83
Successful Mission Performance (%)	54	79	73

Table 1. Comparison of the current Combat Vehicle Crewmen's (CVC) helmet with the new VIS CVC helmet. This shows the reduction in the sound level (SL) at the ear and the increases in the allowable exposure time, speech intelligibility, and mission performance in the Bradley Fighting Vehicle.

between the ear and the ANR microphone (located in the earphone), active attenuation is also reduced.

In another test, a no chin strap condition was added for the CVC because very often tankers unsnap the chin strap. The loss of attenuation is greater for this condition than for the glasses condition with the CVC helmet.

4. CONCLUSIONS

The Army has made tremendous improvements in attenuation and speech intelligibility with the acquisition of the VIS headsets when compared to the older, AN/VIC-1, intercom system with the DH–132 tanker's helmet (Table 1). The mean sound level at the ear minus 1 standard deviation is 99.6 dB(A) for the DH–132 and improves to 83.1 dB(A) for the new CVC. This increases the allowable exposure time from 20 minutes to over 12 hours when evaluated using the 5 dB(A) time-weighted average of the Army's hearing conservation limits. Speech intelligibility increased from 68% to 89%. Garinther and Peters (Ref. 5) demonstrated that crew performance for complicated armor missions is reduced almost linearly as a function of speech intelligibility (Fig. 6). As a result of gains in speech intelligibility with the new VIS intercom and the improved CVC helmet, predictions of successful mission performance increase about 25%.

The addition of the other headsets, such as the CAPS and ACAPS, provides soldiers with hearing protectors that they can use while wearing their combat helmet. Also, when they are with the vehicle, they can be connected to the communications system.

As might be expected, a few problems have arisen during the initial fielding, such as microphone failures and damaged ear seals. The microphone failures have been resolved, and a tougher ear seal is being developed which will better withstand the rigors of the field without affecting attenuation. The soldiers have reported that in spite of the problems, they prefer the comfort of the CVC and the better communications with the new intercom system, with the new headsets, as opposed to the older intercom system. The VIS, with the ANR headsets, has provided soldiers with greatly improved noise attenuation, better communications, and improved capabilities for better performance of its overall mission.

The effectiveness of the ANR headsets has already been demonstrated in the field with the CVC helmet. In the same way the aviation community can profit from the lessons learned by the VIS program and take advantage of tremendous benefits offered by ANR headsets. A well-designed aviator's helmet with ANR can easily achieve, at a minimum, the same attenuation and speech intelligibility in the helicopter cockpit as that obtained by the CVC. From a human factors engineering point of view, the ANR helmet can provide the greatest comfort and ease of donning and doffing required for aviation. Although anecdotal, another benefit reported by the troops is that the perceived level of the workload is reduced. This can be of tremendous benefit in the high workload

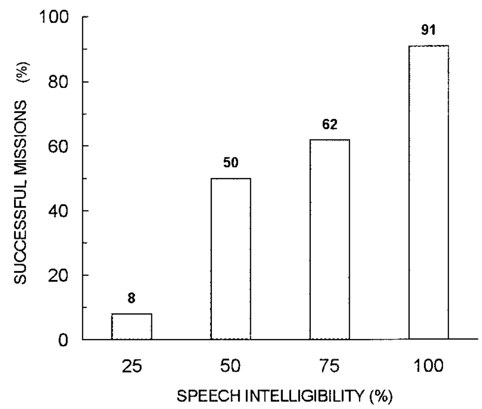


Fig. 6. Armor crew performance as a function of speech intelligibility for complicated mission scenarios (after Garinther and Peters, Ref. 5).

aviation environment. It should be noted that increased discomfort is often perceived as increasing the workload, so it is essential that the headwear is comfortable for the user. Of the many benefits, the most important to the aviation community, as with armor, would be the increase in successful mission performance resulting from improved speech intelligibility. Problems specific to aviation need resolution, but the effort would provide great rewards for the aviation community as it is currently doing for armor.

5. REFERENCES

1. ATCD-CT, "Required Operational Capability (ROC) for the Vehicular Intercommunications System (VIS)," U.S. Army

Training and Doctrine Command, Fort Monroe, Virginia, 1986.

2. Garinther, G.R. and Anderson, B.W., "Enhanced Armor Using the Vehicular Intercommunication System," *Army RD&A*, Sep–Oct 1996, pp. 33-35.

3. Kalb, J.T. and Garinther, G.R., "An Evaluation of Active Noise Reduction when Exposed to the Firing of Paladin," unpublished report, 1995.

4. Dancer, A., Report to the NATO Research Study Group 29 on Impulse Noise, Aberdeen Proving Ground, Maryland, Jul 1996.

5. Garinther, G.R. and Peters, L.J., "Impact of Communications on Armor Crew Performance," *Army RD&A*, Jan–Feb 1990, pp. 1-5.

SPECIAL APPLICATIONS OF ACTIVE NOISE REDUCTION HEADSETS

Richard L. McKinley
Linda J. Morris
Charles W. Nixon

Bioacoustics and Biocommunications Branch
Armstrong Laboratory, AL/CFBA
2610 Seventh Street
Wright-Patterson Air Force Base Ohio 45433-7901, USA

1. SUMMARY

The growth of Active Noise Reduction (ANR) headset technology has accelerated over the past five years. The applications for normal hearing listeners are extensive and the potential for use by persons with hearing loss is excellent. The primary goal of ANR headsets is to reduce the level of the noise at the ears thereby reducing the probability of noise induced effects on hearing and on voice communications. In November 1995, a specially modified ANR headset was demonstrated for users with varying degrees of hearing loss. Most ANR headset systems in operation today are used in aviation associated applications where many of the users have mild to moderate hearing loss. This paper describes the sound attenuation and speech communications performance of both normal and modified ANR headset technology with both normal and hearing impaired users. The limitations and advantages are discussed as well as what can be expected from both standard and modified ANR headset systems.

2. BACKGROUND

Auditory communications remain the vehicle of choice for the rapid and accurate transfer of information. Auditory signals convey information in a wide spectrum of situations that range from the exchange of small talk to a party to notifications of imminent threats and danger in recreational and occupational environments. In the military, and particularly military aviation, a wide variety of auditory signals inform the operator of flight conditions, status of the aircraft and of onboard systems, navigation, and information vital to mission success. The most critical of all of these significant audio signals is voice communications which is indispensable for safety and survival.

The ear is a remarkable mechanism. Its sensitivity enables us to hear a pin drop and its robustness to withstand the intense noise of jet aircraft engines, at least for a time. Military noise environments are equally or more intense than almost all other noise sources. These environments have the unpleasant consequences of voice communications degradation and of temporary and permanent noise induced hearing loss. Temporary hearing losses interfere with the reception of speech and other audio signals during noise exposures and during the time required for the ear to recover after the noise ends. Permanent noise induced hearing loss does not recover to pre-exposure threshold levels and is not responsive to medical treatment.

Persons with noise-induced hearing loss have substantial difficulty with speech recognition in noise. Recognition is impaired by the

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

hearing loss and by the masking effect of the noise on the remaining hearing. Communications are formidable for persons with hearing loss who wear hearing protection devices as well as those using electronically-aided communications systems. The restricted ability of hearing impaired individuals to hear and understand audio signals poses a continuous threat to performance and safety. Job performance is reduced by errors attributed to the inability to hear in the occupational environment. Personal accidents and fatalities are attributed to the employees inability to hear the sounds or warnings associated with the threat or to determine its location in sufficient time to avert the consequences.

In November 1995, Wright-Patterson AFB sponsored an on-base conference and workshop directed to the application of current Air Force laboratory technologies for enabling persons with physical disabilities. The focus of the conference was clearly expressed in its title, Wright Focus On Abilities. The communications were accomplished via seminars, tutorials, and workshops as well as scientific and technical exhibits and demonstrations. The theme was to emphasize and optimize the abilities and skills possessed by each and the potential of new technologies to mitigate the effects of various disabilities. Among the many focus areas addressed were the disadvantages associated with non-normal hearing and deafness.

The celebrity guest and spokesperson for this conference was Miss Heather Whitestone, Miss America 1995. Miss Whitestone has a severe hearing loss in both ears and has virtually no hearing without hearing aids. During coordination of arrangements for her participation in the Wright Focus On Abilities event, Miss Whitestone was invited to take an orientation ride in a high performance fighter aircraft, the F-16.

In order for Miss Whitestone to be able to fly in the two-seat F-16D, she would need to be able to hear and understand voice commands from the pilot in the noisy cockpit. In addition, the flight helmet interfered with the operation of Miss Whitestones's hearing aids causing them to be unusable for this flight. A special helmet was needed to provide Miss Whitestone with communication capability in the F-16 cockpit noise in spite of her severe hearing loss.

3. OBJECTIVE

The objective was to modify a helmet version of an active noise reduction headset to provide sufficient noise reduction and adequate speech level and quality to allow Miss Whitestone to understand voice commands from the pilot in the F-16 cockpit noise under all flight conditions.

4. APPROACH

The three-fold approach was initiated to provide adequate speech intelligibility. First, the noise levels at the ear were reduced by applying passive and active noise reduction. Second, the speech levels were improved by raising the overall gain of the speech signal and then modifying the bandpass of the speech for Miss Whitestone's residual hearing. Third, a highly discriminable closed set vocabulary was developed for the flight.

5. PROCEDURE

The active noise reduction headset was modified by the Bose Corporation to provide gain and bandpass characteristics developed by the authors based on Miss Whitestone's residual hearing. The communications

effectiveness of Miss Whitestone and the F-16 pilot were evaluated with a series of tests in the Biocommunications Laboratory. Performance measurements were made with Miss Whitestone in accurate levels of the F-16 aircraft cockpit noise that would be experienced during the flight. Both the F-16 pilot and Miss Whitestone were trained with the special vocabulary in the F-16 noise environments. Random words and phrases spoken by the F-16 pilot were presented to Miss Whitestone via the modified ANR headset. Miss Whitestone repeated the words and phrases she thought she heard. Her responses were scored by the authors.

6. DATA

The entire 30 item vocabulary was presented to Miss Whitestone three times in the F-16 noise at a level of 105 dB. Miss Whitestone correctly responded to 100 percent of the stimuli presented. This data and other communications performance in noise in the laboratory provided the basis for demonstrating Miss Whitestone's ability to correctly understand voice commands from the F-16 pilot in the F-16 cockpit noise environment.

7. DISCUSSION

The customized helmet/active noise reduction headset system enabled Miss Whitestone to experience a one-hour flight in the F-16 aircraft after which she reported hearing everything perfectly that was said during the flight. Although this concept has been demonstrated in flight only this one time, it does establish that ANR headsets can be customized to significantly enhance the voice communications capabilities of individuals with impaired hearing. This successful effort was accomplished for an individual with a profound hearing loss, one closely approaching deafness. It is expected that customizing the ANR system for the less harsh moderate and moderately severe losses of hearing will be less of a challenge.

It is estimated that the special ANR headset improved the speech-to-noise ratio by 20 to 25 decibels. This increase was obtained from a decrease in noise level at the ear of about 10 to 15 dB combined with an increase in level of the speech signal of up to 10 dB. The quality/fidelity of the high level speech was also improved by the active noise reduction circuitry.

This concept is targeted to individuals with hearing losses that are accompanied by difficulties with voice communications sufficient to keep them from their occupational environments. Military aviators must pass a pure tone audiometric criterion test for retention on flight status. Those who fail the pure tone test because of the amount of the hearing loss can potentially be grounded. These aviators may request a waiver to fly with the hearing loss claiming no communications difficulties in-flight in spite of their pure tone test failure. The moderate hearing losses of many of those who continue to fly can negatively affect the ability to discriminate speech in noise and degrade flight safety. The aircraft pilot is one of the very experienced, highly skilled occupations with exceptionally high investments by their organization in training for the individuals and the motivation to keep them flying is very high. The customized helmet/ANR system should enable a significant number of these highly trained experts to continue or return to flying, with the likelihood that many will experience better voice communications and improved flight safety (via ANR).

The customized ANR system brings those with hearing loss up to communication

performance at normal levels. Noise levels at the ear are reduced. Speech intelligibility is increased with a speech-to-noise ratio gain of up to 20 - 25 dB. The quality of the communication signals are elevated. Subjective increases in comfort and reduced fatigue should also be experienced.

Some basic information about the hearing function of an individual is helpful in customizing the ANR headset. Pure tone audiograms and speech reception thresholds are needed, both aided and unaided. Information on most comfortable and uncomfortable listening level thresholds as well as word recognition scores is also important, preferable for the noise environments in which the headset will be used. Other information, dependent upon the situation, should also be useful in the customizing process.

8. LABORATORY STUDIES

The approach to the full development of the customized ANR headset concept has been extended in scope. A series of laboratory experiments using a large number of subjects with moderate hearing loss has been initiated to systematically examine the amount of improvement in communications that is achievable and the features that are essential to ensure individual success. Three experiments will sequentially investigate the speech intelligibility obtained by normal hearing and hearing impaired subjects while wearing the standard, non-customized ANR headset, and the customized ANR headset with and without additional speech gain.

9. SUMMARY

The customized ANR headset concept has been fully demonstrated only once. It is very promising, relatively easy to fabricate, and cost effective, particularly in terms of the exciting potential benefits. An additional series of laboratory experiments has been initiated to establish small databases of the performance of normal hearing and hearing impaired listeners using the normal passive ANR and customized ANR headsets. Knowledge and experience from these experiments will extend the technology to a larger body of candidates, to additional applications, and to ongoing efforts in both laboratory and inflight scenarios.

ACTIVE NOISE REDUCTION FLIGHT TESTS IN MILITARY HELICOPTERS

CAROL SIMPSON
Psycholinguistic Research Associates
Under contract to US Army Aeroflightdynamics Directorate
Moffett Field, CA 94035-1000, USA
Email: simpson@merlin.arc.nasa.gov

ROBERT KING
Air Operations Division, Aeronautical and Maritime Research Laboratory
Fishermens Bend, Victoria, Australia, 3207
Email: Robert.King@dsto.defence.gov.au

SUMMARY

US Army Aeroflightdynamics Directorate (AFDD), in collaboration with AS Aeronautical and Maritime Research Laboratory (AMRL), has conducted flight tests in a range of military helicopters to determine the potential benefit of active noise reduction (ANR) earcups developed by the UK's Defence Research Agency (DRA) for military aircrew. Test data include (a) acoustic attenuation characteristics, (b) speech intelligibility, (c) aircrew ratings of cockpit speech intelligibility, clarity, and attention demand for speech message recognition, and (d) ratings of the suitability of ANR for operational use. Test aircraft in which data were collected include American NAH-1S (Cobra), UH-1H (Huey), OH-58D (Kiowa), AH-64A (Apache), EH-60 (Blackhawk), and Australian S-70B-2 (Seahawk) and S-70A-9 (Black Hawk). Results show that the DRA ANR system effectively reduced the level of low frequency noise (<800 Hz) and reduced overall at-ear sound pressure levels (SPLs) by around 10 dB. Results also indicate that ANR substantially increases speech intelligibility, reduces the level of attention pilots must use to understand speech communications, works with onboard weapons firing noise, allows pilots to hear familiar audio cues necessary for aircraft situational awareness, and functions without failure in training and actual combat conditions. With the DRA ANR system, speech intelligibility meets the exceptionally high intelligibility criteria as defined in MIL-STD 1472 for operational systems, providing the speech intelligibility needed to ensure that pilots and soldiers communicate tactical information accurately.

1 . INTRODUCTION

There has been a long standing concern with the noise levels experienced by aircrew operating in modern rotary wing military aircraft such as the NAH-1S (Cobra), UH-1H (Huey), OH-58D (Kiowa), AH-64A (Apache), EH-60 (Blackhawk), S-70A-9 (Black Hawk) and the S-70B-2 (Seahawk). At-ear sound pressure levels (i.e., SPLs measured at the ear under the helmet) are high and produce two major operational problems for aircrew. Firstly, aircrew have to limit their exposure times in order to meet current hearing conservation regulations[1] or be provided with additional attenuation devices in order to maintain reasonable manning levels for operational flying [refs 15, 17, 24, 25]. Secondly, high ambient noise levels reduce communications (speech) intelligibility at the ear resulting in reduced operational effectiveness [refs 9, 27].

There are currently three types of device that can be used in conjunction with standard aircrew helmets to provide additional attenuation. These consist of:

(a) soft insert earplugs such as the EAR™ yellow foam earplug[2] which are inserted in the ear canal and worn under the standard helmet,
(b) 'communications' ear plugs (CEPs)[3] such as those developed by the United States Army Aeromedical Research Laboratory (USAARL) which are foam (or triple flange) insert earplugs with a miniature speaker in them, and
(c) active noise reduction (ANR) systems such as that developed by the Defence Research Agency (DRA) which can be fitted in standard earshells and mounted in standard aircrew helmets such as the Advanced Lightweight Protective Helmet for Aircrew (ALPHA), the SPH-4B, the Integrated Helmet and Display Sight System (IHADSS) and the HGU-56 helmet.

It is generally accepted that the use of earplugs in conjunction with the standard aircrew helmet has not proven to be a satisfactory long-term solution to the problems outlined above. While earplugs provide extra attenuation, they also mask inter-communication system (ICS) transmissions further degrading speech

1 Current AS, UK and US hearing conservation guidelines allow a Permissible Daily Exposure Dose (PDED) of 85 dB(A) at-ear for an 8 hour day.

2 The EAR earplug is manufactured by the Cabot Safety Corporation, 5457 West 79th Street, Indianapolis, IN 46268, USA.

3 The CEP is manufactured by Sensor Electronics, Inc., 105 Fairway Terrace, Mt. Laurel, NJ 08054, USA.

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

intelligibility. In addition, earplugs are not universally accepted by aircrew. It has been reported that 20 percent of US Army aviators do not wear ear plugs [ref 19]. Furthermore, they substantially increase helmet donning time.

The approach taken by USAARL for the CEP has been to build upon the dual hearing protection concept. Its earplug component provides additional passive noise attenuation at the ear in the same manner and degree as do current foam insert plugs. It also addresses the problem of reduced speech levels associated with current earplugs. CEPs have an inbuilt miniature speaker which is plugged directly into an adapter on the flight helmet or on the helmet ICS cable for connection to the aircraft ICS system. The speaker faces the ear drum when the CEP is inserted into the ear canal and ICS transmissions are thus not masked by the earplug with this design. All attenuation provided by the CEP is passive. The improvement in signal-to-noise ratio over the current dual protection system is obtained by increasing the speech level received at the ear drum while maintaining the passive attenuation provided by the ear plug.

The objective of ANR is to both increase mission effectiveness of the aircrew and to eliminate the need for dual hearing protection devices, i.e. earplugs in addition to flight helmet. ANR improves the signal-to-noise ratio for speech and other intercom signals thereby improving speech intelligibility of incoming radio transmissions and of intercom transmissions for the aircrew; the detectability and recognition of certain other acoustic cues is also improved. There is a corresponding reduction in the attention demand required for the crew to accomplish their communications tasks. In theory, the achievement of any of these specific performance improvements acts as multiplier to increase the mission effectiveness of the aircrew. In combination, these improvements are expected to have an even greater positive effect on mission effectiveness. In fact, improvements in intelligibility were found to correlate with improved mission task performance and overall mission performance by armor crew in a series of experiments conducted by US Human Research and Engineering Division (HRED) of the Army Research Laboratory (ARL) [ref 27]. By extension, a similar beneficial effect is expected for aircrew.

Unlike the passive noise attenuation provided by earplugs, ANR actually cancels some of the noise by generating an acoustic waveform that is (ideally) 180° out of phase with the noise inside the earshell and adding this 'anti-noise' to the earshell. The exact specifications of particular ANR systems vary in frequency range, amount and spectral composition of attenuation provided, system weight, power consumption, and other parameters.

The Aerospace Division of DRA has been developing an ANR system mounted within the earshells of standard aircrew helmets for a number of years. The pre-production prototype ANR circuit adds a net of 21 grams to each earcup. In the DRA ANR system, noise in the earshell is sampled via a microphone, phase inverted and then reintroduced into the earshell. A separate circuit pre-emphasizes ICS audio, using a filter with characteristics that are the inverse of the active attenuation effect of the ANR circuit, and reintroduces it directly into the earshell so as to preserve the original level and spectral characteristics of the ICS audio. Communications integrity is ensured by means of a fail safe circuit should the ANR lose power or fail to operate. When the ANR is off or has insufficient power, the ICS audio bypasses the ANR circuit so that communications capability remains even if the ANR system fails.

The AeroFlightDynamics Directorate (AFDD) and the Aeronautical and Maritime Research Laboratory (AMRL) have extensively investigated the performance of the DRA ANR system. The acoustic performance of the DRA ANR system has been measured in-flight in the Australian S-70B-2 (Seahawk) and S-70A-9 (Black Hawk) aircraft. Speech intelligibility scores and aircrew ratings of intelligibility, clarity, attention demand for speech message recognition and the suitability of each system for operational use have also been collected in-flight in the American NAH-1S (Cobra), UH-1H (Huey), OH-58D (Kiowa), AH-64A (Apache), EH-60 (Blackhawk), and the Australian S-70B-2 (Seahawk) and S-70A-9 (Black Hawk). The aim of the present paper is to report, examine and discuss representative data from these trials.

Work in progress is also examining the performance of the USAARL CEPs and the Bose ANR system. The attenuation characteristics of the CEP system have been measured in-flight in the S-70B-2 (Seahawk) aircraft and aircrew rating data for the CEP and Bose ANR systems has been collected in-flight in the NAH-1S (Cobra). Preliminary results are reported.

2. EQUIPMENT AND EXPERIMENTAL PROCEDURE

2.1 ANR System

The DRA ANR system was supplied by the Defence Research Agency (DRA). The Bose ANR system and

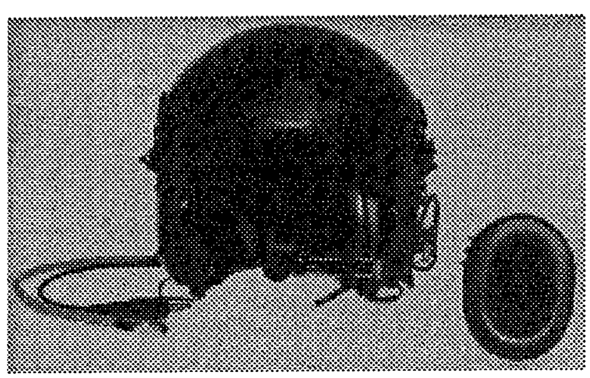


Figure 1: The DRA ANR system. The ANR system is mounted in a standard earshell and fitted to individual aircrew helmets.

CEPs were supplied by the Bose Corporation and USAARL respectively.

2.2 Aircraft

Acoustic measurements of DRA ANR and CEP systems were made in-flight in the S-70A-9 (Black Hawk) and/or the S-70B-2 (Seahawk) aircraft. Speech intelligibility scores and aircrew ratings of intelligibility, clarity, attention demand for speech message recognition, and system suitability for operational use were collected in-flight for the DRA ANR system in the American NAH-1S (Cobra), UH-1H (Huey), OH-58D (Kiowa), AH-64A (Apache), EH-60 (Blackhawk), and Australian S-70B-2 (Seahawk) and S-70A-9 (Blackhawk). Intelligibility and rating data were collected in a comparative study of three systems, DRA ANR in SPH-4B earcups, Bose ANR in an HGU type earcup, and the triple flange plug version of the CEP, in the American NAH-1S (Cobra).

2.3 Acoustic Measurement and Analysis Equipment

Acoustic data were measured using the Head Acoustic Measurement System (HAMS; see Figure 2). The HAMS is a dummy head device which has been specifically designed for noise measurement and meets IEC 711 standards for coupled measurements (i.e., measurements where the ear canal is coupled to a headset or helmet [ref 14]). The HAMS represents a considerable advance over traditional acoustic measurement and recording devices such as stand-alone microphones and sound level meters because it incorporates the representative acoustic transfer characteristics of the human head and torso when measuring the acoustic environment [ref 20].

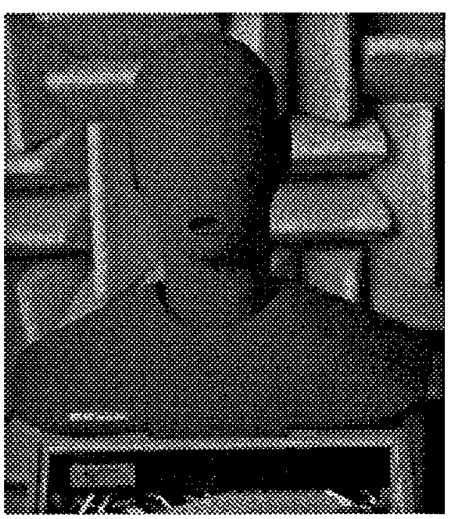


Figure 2: The Head Acoustic Measurement System

The HAMS consists of a dummy head, ear canal simulators and a 'torso box' containing equalization and recording equipment. Left and right outputs from microphones in the ear canal simulators are equalized and passed to a digital audio tape recorder with a sampling frequency of 44.1 kHz. The microphones have a linear frequency response that is effectively flat between 20 Hz and 20 kHz (± 1dB). Each channel was calibrated using a Brüel and Kjær 4230 sound level calibrator. Recordings were analysed using a Hewlett-Packard 3567A dual channel spectral analyser.

The HAMS was 'fitted' with a Mk IV ALPHA (Advanced Lightweight Protective Helmet for Aircrew) for the acoustic measurements. For the DRA ANR measurements, standard earshells incorporating the DRA ANR system were fitted to the ALPHA helmet. For the CEP measurements, the CEPs were inserted in the HAMs ear canals under the ALPHA helmet.

2.4 Acoustic Analysis Procedure.

Third octave band analyses were performed in order to determine the acoustic attenuation characteristics of the DRA ANR and CEP systems. For the ANR system, attenuation was defined as the difference between SPLs measured in each $^1/_3$ octave band with (a) ANR On, and (b) ANR Off in each flight condition and measurement position. Measurements with ANR On and ANR Off were repeated four times with the ALPHA helmet being removed and refitted to the head for each measurement. Attenuation performance was measured at a number of crew positions in the S-70B-2 Seahawk (Pilot and Sensor operator positions — see Figure 3) and S-70A-9 Black Hawk helicopters (Pilot, Loadmaster, Middle and Rear positions — see Figure 3) under a variety of flight conditions (hover, transition, cruise, deceleration with doors open and shut). The mean attenuation provided by the DRA ANR system in each $^1/_3$ octave band and its associated standard deviation is reported.

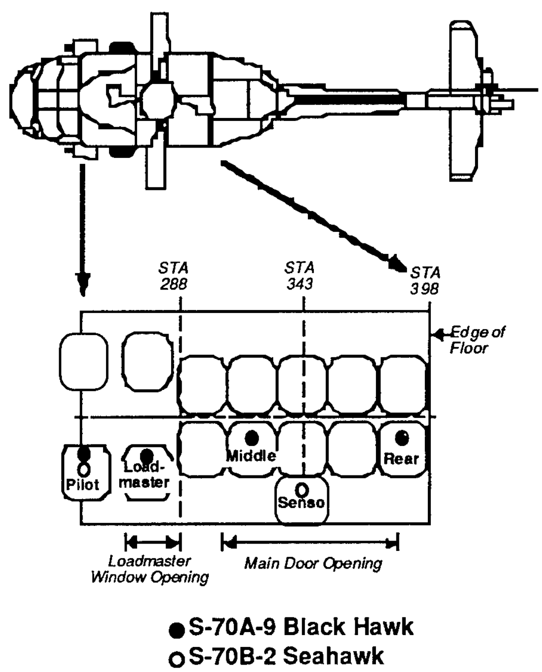


Figure 3: Acoustic measurement positions in the S-70A-9 (Black Hawk) and S-70B-2 (Seahawk) helicopters.

For the CEP system, attenuation was defined as the difference between SPLs measured in each $^1/_3$ octave band with (a) the earplug inserted in the ear canal under the ALPHA helmet, and (b) without the earplug inserted in the ear canal under the ALPHA helmet in each flight condition and measurement position. Measurements with and without the earplug inserted were repeated four times (with the earplug and/or the ALPHA helmet being removed and refitted to the head for each measurement). Attenuation performance was measured at the Senso position in the S-70B-2 (Seahawk) helicopter during cruise flight with doors shut. The mean attenuation provided by the CEP system in each $^1/_3$ octave band and its associated standard deviation is reported.

2.5 Attenuation Performance of the DRA ANR and CEP Systems in a Typical Military Rotary Wing Aircraft

The acoustic characteristics of the noise in the S-70A-9 Black Hawk, and the 'conservatively adjusted' attenuation characteristics[4] of the properties of the ALPHA helmet (based on measurements at a number of crew positions in the S-70B-2 and S-70A-9 Black Hawk helicopters under a variety of flight conditions) have been reported previously [refs 15, 17].

In order to provide a representative comparison of the acoustic attenuation performance of the DRA ANR and CEP systems, the performance of each system in a typical military rotary wing aircraft (the S-70A-9 Black Hawk) was modelled. Performance was modelled by calculating the conservatively adjusted at-ear SPLs that would be experienced by aircrew wearing an ALPHA helmet with each system at the Pilot, Loadmaster, Middle and Rear positions in the S-70A-9 (Black Hawk) during cruise flight (see Figure 3). At-ear SPLs were calculated by subtracting the 'conservatively adjusted' attenuation provided in each $^1/_3$ octave band by each system (as reported in this paper) from the conservatively adjusted at-ear SPL in each $^1/_3$ octave band experienced by aircrew wearing the standard ALPHA helmet at these position during cruise flight [as reported in ref 15]. The resultant $^1/_3$ octave spectra were A-weighted and integrated across bands to provide the A-weighted Overall SPL (OASPL) at-ear with ANR and CEP. The 'conservatively adjusted' at-ear SPLs experienced by aircrew wearing the 'stock' ALPHA helmet in at each position are also reported so that the reader can gauge the effectiveness of the DRA ANR and CEP systems.

2.6 Speech Intelligibility

The speech intelligibility flight test approach combines a traditional seven-point numeric rating scale method [ref 11] with traditional speech intelligibility methodology [ref 2]. The rating scales were validated by comparing aircrew ratings of speech intelligibility and other speech features to their actual listening performance as measured by a Phonetically Balanced (PB) Word Test. The PB Word test was selected for this flight testing program, in preference to the Modified Rhyme Test [refs 8,13] or the Diagnostic Rhyme Test [ref 26] because it is the most sensitive to small differences in speech intelligibility and is also the only one that tests all the phonemes of English. Three rating scales for 1) speech intelligibility, 2) speech clarity, and 3) attention demand were developed and administered in accordance with standard psychological test methodology, and the resulting rating data were compared to PB Word test results, within subjects. Chi-square tests were performed to test for correlation of each type of rating scale data with the PB Word Test data. Data obtained with operational pilots for each of the three rating scales were consistently found to be correlated to PB word intelligibility data obtained from these same pilots. Statistical results are presented with each of the four flight tests reported below. Significance levels of the correlation between pilots' performance and their ratings ranged from $p<0.01$ to $p<0.001$.

The AFDD Flying Laboratory for Integrated Test and Evaluation (FLITE) NAH-1S Cobra helicopter was used as a testbed to 1) develop and test a portable speech intelligibility test equipment package that could be used safely and efficiently in flight, 2) test the standard PB-word test procedures and methodology and adapt them to the constraints of the helicopter flight environment, 3) develop and test the set of speech rating scales.

A small flight test package was developed consisting of a tape player, a custom amplifier, and an audio switching cable with the entire package mounted on a standard flight kneepad. All components of the test package are worn in or on the flight clothing of the pilot or the experimenter and are battery powered to eliminate any need for aircraft electrical power. PB Word lists in random orders are pre-recorded on audio tape at controlled levels, calibrated to a 1000 Hz tone, also recorded on the tape. Audio output from the tape recorder is introduced into the front seat ICS microphone input via a Y-cord with a two-position switch. In the normal position, the audio signal from the experimenter's flight helmet microphone is the only signal introduced into the front seat ICS microphone input. In the tape position, the tape recorder output is the only signal introduced into the front seat ICS microphone input. In this way, the experimenter can either talk to the pilot or can present test speech tokens to the pilot using the normal ICS audio system. This ensures that the test tokens are presented using the actual operational audio system characteristics of the helicopter. Since current Army ICS systems are limited in bandwidth, and this bandwidth restriction reduces speech intelligibility, it is critical to a test of operational speech intelligibility to use the actual aircraft ICS system.

4 The attenuation factor is conservatively adjusted by subtracting one standard deviation from the mean attenuation in each $^1/_3$ octave band. This adjustment ensures that the reported degree of noise reduction would be obtained on 80% of occasions[ref 1].

Test procedures for administering the PB Word Test and the rating scales were adapted for application to pilots who are actively hand-flying the aircraft and performing training and/or actual missions during data collection. The adaptations provide one or more benefits in comparison to traditional laboratory procedures. All test stimuli are pre-recorded on audio tape and are presented aurally to the pilot. The pilot providing the data gives all responses verbally rather than in written form. The responses are recorded on audio tape and written by the experimenter as they are spoken by the pilot. This frees the pilots hands and eyes for the primary task of flying the aircraft. To preclude the experimenter mis-hearing the pilot's response when those responses consist of a PB test word the pilot heard, the pilot says the word itself and then uses it in a short phrase or spells it via the phonetic alphabet, whichever is easier for each pilot. To ensure that the pilot remembers the end points of each rating scale, the experimenter says these each time a rating is requested. The experimenter(s) conducting the data collection in the cockpit are not only trained in experimental psychology but are also pilots. They are therefore aware of safety of flight issues and can take precautions against the testing interfering with safety of flight.

The Central Institute for the Deaf (CID) W-22 phonetically balanced (PB) word lists [ref 18] were chosen for these studies. The CID lists are superior to the Harvard PB Word Lists [ref 7] in that the phonemic and phonetic balance of phonemes within lists is better within and across the W-22 lists than for the Harvard lists.

During initial flight testing with the W-22 PB word lists, it became apparent that the flight time consumed by administering the several different 50-word lists needed for a controlled study was long as well as fatiguing for both the pilot and the Experimenter. Accordingly, a set of 25-word lists, developed and validated by Campbell (1965), were used instead [ref 4]. These 25-word lists are composed of the same set of words as the CID W-22 50-word lists. In composing these lists, Campbell used word difficulty data obtained from a pool of military veterans who had received audiological testing for measurement of speech discrimination losses. He found that average percent error rate for the 200 different words in the CID test set ranged from 0% to 86% and that the average error rate for each of the four CID W-22 lists ranged from 22% to 26%. By reassigning words to half-lists of 25 words each, Campbell reduced the variation in average error rate from four percentage points to one percentage point across half-lists. By using the Campbell 25-word lists, it was possible to both reduce the test time per list and to make finer discriminations in intelligibility.

When the 25-word lists were used in the FLITE Cobra, the pilots and the Experimenter reported less fatigue, and list presentation time was reduced by 50%, making possible the collection of speech intelligibility data for an 8-cell experimental matrix within a 48 minute flight period, not including take-off, flight to and from the test area, and landing.

In addition to the modified procedure for collecting PB word intelligibility scores, a set of rating scales was developed and tested. Rating scales have the advantage of ease and speed of administration in comparison to the presentation of a list of words for identification.

However, it was necessary to determine whether pilots could easily and reliably make ratings of speech quality and whether their ratings would correspond to the intelligibility scores obtained by the PB word intelligibility method. Three seven-point scales were designed to measure pilots' ratings of 1) perceived speech intelligibility, 2) perceived clarity of the speech, and 3) perceived attention demand to recognize the words. The use of ratings scales is not new to the evaluation of aircraft systems. The Cooper Harper Handling Qualities Ratings which have long been used to measure the flight handling characteristics of aircraft are basically a set of rating scales [ref 5].

One or two 25-word PB Lists are used per cell in the experiment matrix with a different assignment of lists to cells for each pilot. Each list requires 6 minutes of flight time. For each test condition the appropriate listening level is set, based on peak word linear sound pressure level (SPL) measured at the pilot's ear canal entrance via a miniature probe microphone which does not break the seal on the helmet earcup. The 25-word PB list is then presented to the pilot, one word at a time. The pilot speaks his response, consisting of the word he has perceived followed by that word used in a short phrase or spelled using the aviation phonetic alphabet, at his option. The experimenter then reads back to the pilot what the experimenter has heard as the pilot's response. Then the experimenter plays the next word on the list to the pilot. At the end of each list, the pilot is instructed to give his speech ratings of intelligibility, clarity, and attention demand for that list of words. All voice communications between pilot and experimenter are recorded on a second, portable, battery-operated tape recorder.

The requirements for military communications systems intelligibility in MIL-STD-1472 were the standard against which PB word intelligibility was compared.

2.7 Suitability for Military Flight Operations.

At the end of each flight, each aircrew member also completed a questionnaire composed of Yes/No questions and rating scales to collect data on the consistency, stability, reliability, and operational suitability of each helmet configuration over the course of the flight[5]. In this questionnaire aircrew were also asked to give overall ratings of the speech just heard for each helmet configuration using the three seven-point scales:

---

5 For a copy of the questionnaire, contact the first author

(a) speech clarity (with 1 'low' and 7 'high'),
(b) speech intelligibility (with 1 'low' and 7 'high'), and
(c) the attention demand required for understanding speech (with 1 'low attention needed' and 7 'high attention needed').

2.8 Comparative Flight Tests.

Once the test procedures had been developed and tested in the FLITE Cobra, flight tests were conducted over the course of seven years in several military helicopters including but not limited to American OH-58D, AH-64A, NAH-1S, AH-1F, UH-1H, UH-60A, EH-60, CH-47, and Australian S-70B-2 (Seahawk) and S-70A-9 (Black Hawk). These tests were conducted with operational pilots in the field in order to provide a more severe test of the system in a realistic environment. Controlled studies of speech intelligibility were conducted in four individual studies, each using the same PB Word and rating scale methodology described above. Additionally, after the rating scales had been validated against PB Word intelligibility, rating scale data were collected from pilots as they performed training missions in the US and actual combat missions during operations Desert Shield and Desert Storm. Three of the controlled studies examined performance and ratings with DRA ANR in comparison to the aircrew's stock flight helmets. The fourth study included the USAARL CEP and a US-developed ANR system as additional experimental systems.

Study 1 - OH-58D (Kiowa)

The first field test of speech intelligibility was conducted during a two-week flight test of ANR in the field in the OH-58D with the pilots of Alpha Co., 3/24 AVN BGDE, at the Ft. Stewart Army Reservation, Hunter AAF, Georgia, during brigade field exercises, May 30 to June 8, 1989. A detailed description of this experiment can be found in Simpson and Gardner, 1991 [ref 21]. The ANR was installed in pilots' stock OH-58D (Advanced Helicopter Improvement Plan - AHIP) TEMPEST-qualified flight helmets and flown NOE during day and night reconnaissance missions. Linear sound pressure levels, measured inside the flight helmet earcups at the pilots' ears, were reduced overall from 10 to 20 dB compared to the stock helmet, speech communications were clearer and more intelligible for the ANR modified helmets compared to the pilots' stock helmets, and the pilots' ratings also indicated that they required less attention to understand the speech when using ANR. The OH-58D mission is highly dependent on effective tactical communications for coordination of both scout and attack missions. Rapid, accurate information transfer is critical to timely detection of the enemy position and to effective, coordinated attack on threats and targets alike.

The baseline helmet for comparison in this test was the standard TEMPEST OH-58D SPH-4A (AHIP) flight helmet without ANR installed. For each pilot, communications effectiveness, as measured by PB Word intelligibility and by speech ratings, was compared for three conditions: 1) his stock helmet, 2) the same helmet with DRA ANR Mark IV earcups with ANR turned off, and the same helmet with DRA ANR Mark IV earcups with ANR turned on.

Data were collected by a team composed of an OH-58D pilot in the right seat and the experimenter in the left seat. The Experimenter used a standard SPH-4 helmet with no modifications and monitored appropriate radio frequencies as directed by the pilot. The pilot wore either his stock OH-58D helmet or his stock OH-58D helmet with ANR ear cups installed, depending on the test condition.

Speech Intelligibility Measurement

PB Word speech intelligibility and speech ratings were measured with cockpit doors off and heater, blower, and mast-mounted sight (MMS) also off. Two levels of flight task difficulty were used — a baseline level and a difficult level. For the baseline level the helicopter was sitting on the ground with engines running, rotors engaged, flat pitch, 100% RPM. The difficult level was nap-of-the-earth (NOE) flight. In addition, two listening levels were used: preferred level and reduced level. For Pilot's Preferred Listening Level, the pilot selected the ICS volume at which he wanted to hear the words. The reduced level was used to simulate weak transmissions. The speech in this case was at a volume about 3 dB below the pilot's preferred level (see Table 1).

Table 1. Test Condition Matrix For OH-58D.

HELMET CONFIGURATION

Stock SPH-4A (no ANR)	SPH-4A ANR Off	SPH-4A ANR On

LISTENING LEVEL x DIFFICULTY

Pilot's preferred level x Baseline
3 dB below preferred level x Baseline
Pilot's preferred level x Difficult
3 dB below preferred level x Difficult

TEST CONDITION MATRIX FOR OH-58D
Baseline = Flat pitch, on ground, 100% RPM
Difficult = NOE flight

Order of helmet condition was balanced across pilots. For a given helmet condition, each pilot listened first with his preferred listening level and then with the reduced level using a different word list. The baseline flight level was always run first, followed by hot refuel (i.e., engine running during fueling), followed by the difficult level of NOE flight. Thus there was a potential confounding of listening level with practice and of

flight difficulty level with practice. Safety reasons dictated the fixed order of flight difficulty while the need to first determine each pilot's preferred level as his personal baseline led to the fixed order of listening level.

Flight Tasks During Intelligibility Testing

The actual flight configurations and maneuvers that the pilots flew during collection of intelligibility data were coordinated with the Principal Pilot, the four Participating Pilots, Safety Officer, and Operations Officer of A CO, 3/24th AVN REGT. Segments of low level, contour, and NOE were flown as appropriate and as agreed by the participating pilots. The choice of maneuvers was made after the pilots had become familiar with the Speech Intelligibility Testing Procedures. Each participating pilot performed the intelligibility tests for each of the helmet configurations, for the two difficulty levels of flight.

Post-flight Questionnaires

Pilots also completed a post-flight questionnaire on Active Noise Reduction and on Speech Intelligibility and Communications Task Performance. The questionnaire contained rating scales and yes/no questions on the stability, comfort, and operational suitability of ANR.

Study 2 - Australian S-70B-2 (Seahawk)

Study 2, conducted in April 1992, was similar in design to Study 1 except that the test aircraft was an Australian S-70B-2 (Seahawk) and the aircrew were Royal Australian Navy pilots and sensor operators. The flight helmets used were the Alpha helmets which are stock for aircrew flying this aircraft. Two pilots and two sensor operators of Tiger Squadron 816, Seahawk Introduction and Transition Unit at Naval Air Station Nowra, Australia, participated. DRA ANR Mark IV Mod 1 in Alpha Mark IV earcups were fitted to the Alpha helmets and data were collected for ANR On and ANR Off conditions. The same test equipment package and procedures were used as had been used in Study 1 in the OH-58D. PB word data, speech rating data, and operational suitability data were collected. Scoring of the PB word data was corrected for the Australian accent and associated differences in speech perception for the Australian aircrew as compared to American aircrew [ref 23].

Study 3 - American EH-60 (Quickfix)

Study 3 was conducted in October 1992 in American EH-60 aircraft with pilots and operators from Quickfix Platoon, A Company, 3/24th Aviation Brigade. The same test equipment package and procedures were used as had been used in Studies 1 and 2 except that for this study PB test words were transmitted via radio from a portable ground unit to the aircraft in flight. Four aircrew wore SPH-4 flight helmets with thermal plastic liners installed and fitted with DRA Mark IV Mod 1 ANR in Alpha Mark IV earcups. PB word data, speech rating data, and operational suitability data were collected [ref 10].

Study 4 - American NAH-1S (Cobra)

Study 4 was conducted in the AFDD FLITE NAH-1S Cobra in July-September 1993 and was designed to compare ANR and the CEP for speech intelligibility and operational suitability. Six Cobra pilots from the Air National Guard at Stockton, California each flew four 2-hour flights with each of four helmet configurations:

1) Stock HGU-56/P[6], 2) HGU-56/P with the triple flange CEP worn under the stock earcups, 3) HGU-56/P with DRA SPH-4B Mod 2 ANR earcups, and 4) HGU-56/P with custom ANR earcups by Bose Corporation[7]. The same test equipment package and procedures were used as had been used in Studies 1 and 2. As with Study 3, PB test words were transmitted via radio from a portable ground unit to the aircraft in flight. PB word data, speech rating data, and operational suitability data were collected [ref 22].

3. RESULTS

Results and statistical analyses for all studies are presented next. Interpretation of the results will be covered in the Discussion section.

3.1 Attenuation Characteristics of the DRA ANR and USAARL CEP Systems.

Third octave band analyses were performed in order to determine the attenuation characteristics of the DRA ANR and CEP systems. Attenuation was defined as the difference between SPLs measured in each $^1/_3$ octave band with ANR On and ANR Off, or with the CEP inserted or not inserted under the helmet, respectively. Measurements were repeated four times and the mean attenuation provided by each system in each $^1/_3$ octave band calculated. The mean attenuation and the standard deviation associated with it is shown for each system in Figure 4.

The DRA ANR system provides a substantial level of attenuation ($\geq$ 10 dB) in $^1/_3$ octave bands centered between 50 and 400 Hz, and some attenuation (5 to 10 dB) in $^1/_3$ octave bands centered between 500 and 800 Hz. The DRA ANR system also produces some amplification (2 to 8 dB), occurring in $^1/_3$ octave bands centered between 1000 and 1600 Hz. The CEP system provides a substantial level of attenuation ($\geq$ 10 dB) in $^1/_3$ octave bands centered between 50 and 315 Hz, and some attenuation (5 to 10 dB) in $^1/_3$ octave bands centered between 400 and 630 Hz. The CEP provides excellent attenuation (> 20 dB) in $^1/_3$ octave bands centered between 2 kHz and 10 kHz. Overall, it can be seen that the DRA ANR generally provides greater attenuation at

6 The HGU-56/P (prototype) flight helmets were supplied by Project Manager for Air Crew Life Support Equipment (ALSE) through collaboration with Mr. Ben Mozo at USAARL.

7 Bose Corporation loaned their ANR system to AFDD for this study.

low frequencies, particularly in $^1/_3$ octave bands centered from 80 to 160 Hz and 250 to 630 Hz. The CEP system generally provides greater attenuation at higher frequencies, particularly in $^1/_3$ octave bands centered from 800 Hz to 10 kHz.

Figure 5 shows the 'conservatively adjusted' (mean minus one standard deviation) attenuation afforded in each $^1/_3$ octave band by (a) the ALPHA helmet, (b) the ALPHA helmet and the DRA ANR system, and (c) the ALPHA helmet and the CEP system. The ALPHA helmet provides good attenuation (>15 dB) in $^1/_3$ octave bands centered between 315 and 10000 Hz, and some attenuation (5 to 10 dB) in $^1/_3$ octave bands centered between 200 and 250 Hz. Noise levels in bands centered between 50 and 160 Hz are amplified under the ALPHA helmet, suggesting some resonance occurs in the ear canal, or the helmet transfers some low frequency noise via bone conduction. When viewed as an integrated unit, the ALPHA helmet fitted with the DRA ANR system provides effective broadband attenuation. The DRA ANR system provides good attenuation at the lower frequencies where the ALPHA helmet (without ANR fitted) produces slight amplification (50 to 160 Hz), while the ALPHA helmet (without ANR fitted) provides

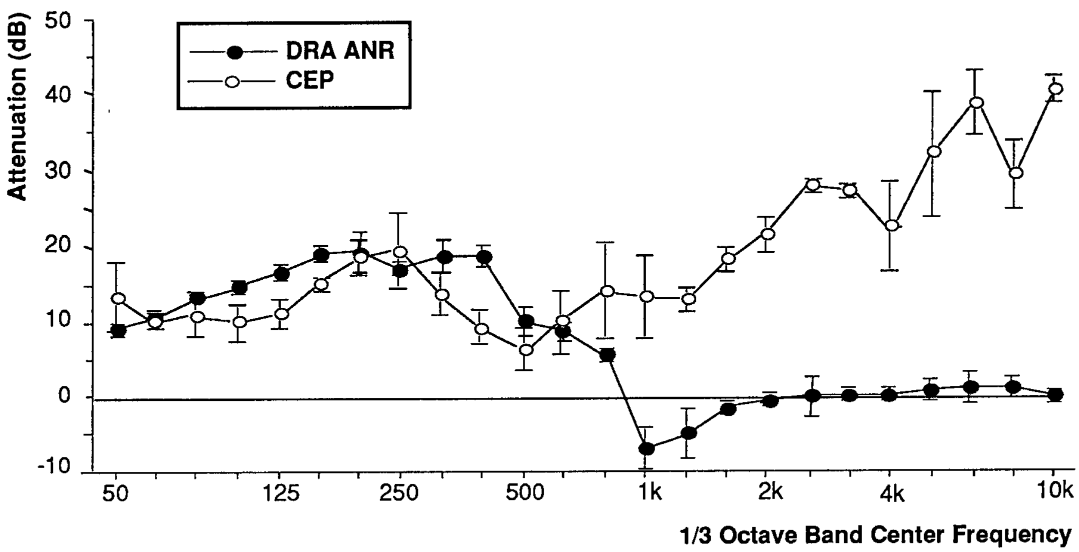


Figure 4: Mean and standard deviation, one-third octave band attenuation characteristics of the DRA ANR and CEP systems.

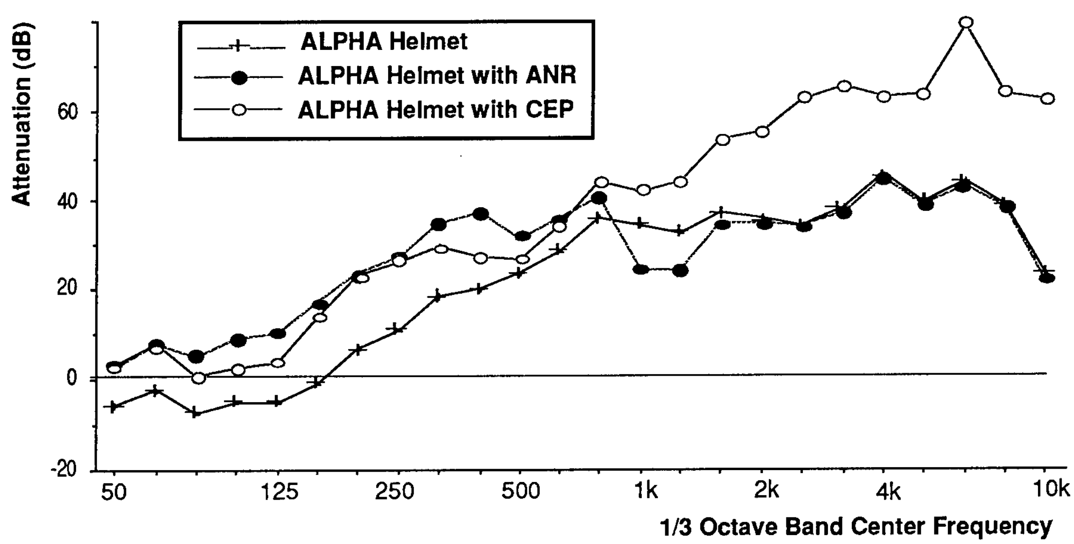


Figure 5: Conservatively adjusted one-third octave band attenuation characteristics of the (a) ALPHA helmet, (b) ALPHA helmet with DRA ANR system, and (c) ALPHA helmet with CEP system.

adequate passive attenuation at the frequencies where the integrated ANR system produces some amplification. As figure 5 shows, the CEP system provides additional attenuation at all frequencies. Low frequency (<800 Hz) attenuation gain is generally not as great as that seen with ANR. High frequency (>800 Hz) attenuation gain is greater than that seen with ANR.

3.2 Performance of the DRA ANR and CEP Systems in a Typical Military Rotary Wing Aircraft

In order to provide a representative comparison of the acoustic performance of the DRA ANR and CEP systems, the performance of each system was modelled in a typical military rotary aircraft, the S-70A-9 Black Hawk. Performance was modelled by calculating the overall conservatively adjusted at-ear SPLs that would be experienced by aircrew using each system at the Pilot, Loadmaster, Middle and Rear positions during cruise flight in the Black hawk. These overall SPLs were obtained by subtracting the conservatively adjusted attenuation provided in each $^1/_3$ octave band by each system (as reported above) from the conservatively adjusted at-ear SPL experienced by aircrew wearing the standard ALPHA helmet in this flight condition [as reported in ref 15]. Table 2 shows:

(a) the ambient cabin noise level at the Pilot, Loadmaster, Middle and Rear positions during cruise flight in the S-70A-9,
(b) the at-ear SPL that would be experienced by aircrew wearing the ALPHA helmet only,
(c) the at-ear SPL that would be experienced by aircrew wearing the ALPHA helmet fitted with the DRA ANR system, and
(d) the at-ear SPL that would be experienced by aircrew wearing the ALPHA helmet in conjunction with the CEP system.

As Table 2 shows, high cabin noise levels are generated in the S-70A-9 during cruise flight, with levels between 106.4 dB(C) and 108.9 dB(C) occurring at the Pilot, Loadmaster, Middle and Rear positions. The ALPHA helmet has good passive attenuation properties and serves to reduce the cabin noise levels to 90.3 dB(A), 86.5 dB(A), 87.4 dB(A) and 88.1 dB(A) at-ear at the Pilot, Loadmaster, Middle and Rear positions respectively. Using the DRA ANR system with the ALPHA helmet would see at-ear SPLs further reduced to 79.9 dB(A), 77.7 dB(A), 77.7 dB(A) and 77.1 dB(A) at these positions, while using the CEP system in conjunction with the ALPHA helmet would reduce at-ear SPLs at these positions to 81.6 dB(A), 78.6 dB(A), 77.7 dB(A) and 77.8 dB(A) respectively. When the ALPHA only levels are compared to the levels obtained with DRA ANR and CEP, respectively, the mean overall additional overall attenuation provided by the DRA ANR system was 10.0 dB, while the mean overall attenuation provided by the CEP system was 9.2 dB.

3.3 Speech Intelligibility: Performance and Ratings

Figure 6 shows the PB Word intelligibility data for Studies 1, 2, and 3 in the OH-58D, S-70B-2, and EH-60, respectively. The data are means for each of the three aircraft. Data for four aircrew are averaged within each of the aircraft; altogether there were 12 aircrew (4 per study x 3 aircraft) across all three studies. In addition, on the far right side, is shown the overall mean across all three aircraft and the 12 aircrew. For comparison the "exceptionally high" and "normally acceptable" PB Word score criteria for operational systems as defined in MIL-STD-1472 are shown as dotted lines at 90% and 75% PB word scores, respectively.

For each of the three studies, the PB Word scores were higher with ANR ON than with the stock helmet configuration. These differences were all statistically significant as tested by Wilcoxon's Signed Ranks Test for Matched Pairs. The pairs of data consisted of Stock Helmet versus ANR ON for otherwise identical listening conditions for each of the 4 pilots. For conservatism, even though we predicted the direction of the experimental effect and could have justifiably used a one-tailed test, all tests for significance were conducted using a two-tailed test.

For Study 1 in the OH-58D the independent variables were, in addition to helmet condition: Two Listening Levels and Two Levels of Flight Task Difficulty. There were thus 16 matched pairs of data. The positive effect on PB word intelligibility for ANR ON compared to the Stock helmet was highly significant ($n=16$, $R=10$, $p<.01$, two-tailed test). The significant effect on intelligibility is impressive given that these data were collected in the field with actual users in the presence of numerous sources of experimental "noise", i.e.,

Table 2. Ambient and conservatively adjusted at-ear SPLs in the S-70A-9 helicopter at the Pilot, Loadmaster, Middle and Rear positions during cruise flight.

	Pilot Position	Ldmaster Position	Middle Position	Rear Position
Ambient Cabin Noise Level	107.9 dB(C)	106.4 dB(C)	108.9 dB(C)	107.6 dB(C)
At-ear SPL, ALPHA only	90.3 dB(A)	86.5 dB(A)	87.4 dB(A)	88.1 dB(A)
At-Ear SPL, ALPHA with DRA ANR	79.9 dB(A)	77.7 dB(A)	77.7 dB(A)	77.1 dB(A)
At-Ear SPL, ALPHA with CEP	81.6 dB(A)	78.6 dB(A)	77.7 dB(A)	77.8 dB(A)

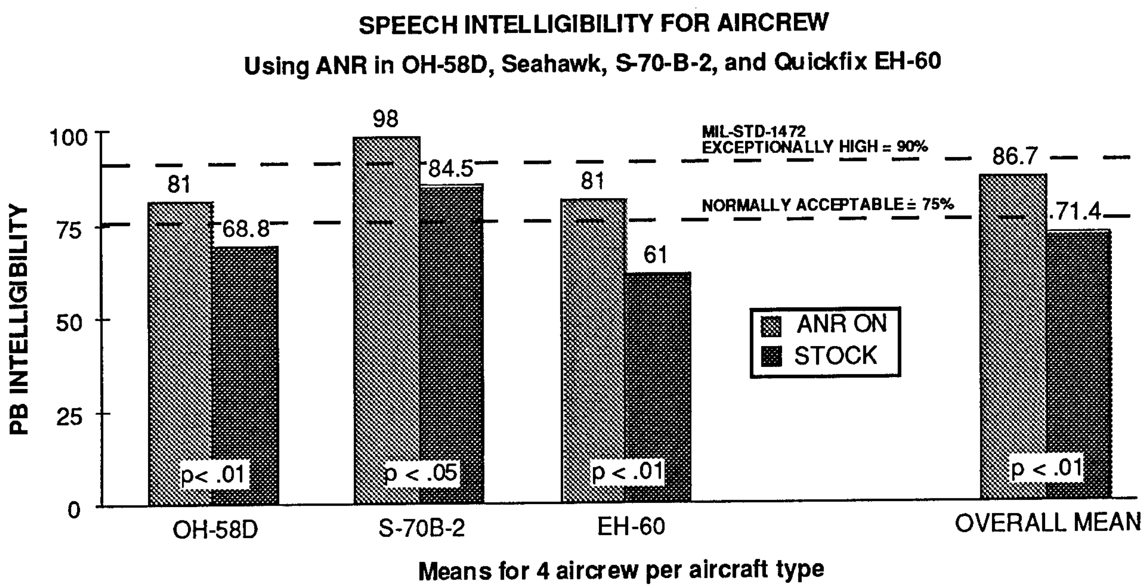


Figure 6: Phonetically Balanced (PB) word intelligibility for three aircraft for stock helmet compared to helmet with ANR ON.

variability in factors which cannot be controlled in the field the way they could be in the laboratory. Age of the four OH-58D pilots ranged from 26 to 42 years. Their hearing ranged from no loss to one pilot who was flying on a hearing waiver due to high frequency hearing loss. In addition, one pilot did not have a good fit of his ANR helmet but neglected to inform the experimenter until after the testing. This poor fit intermittently degraded attenuation and speech intelligibility for that pilot for the ANR helmet regardless of whether ANR was ON or OFF. Our test of statistical significance was performed using all data pairs for all pilots.

Figures 7, 8, and 9 show the average results for each of the three speech rating scales: intelligibility, clarity, and attention demand, respectively. The format is the same as for the PB Word data, showing mean data for each of the three aircraft and an overall mean across all three aircraft and all twelve aircrew. The same test for significance was applied to the pilots' ratings for: 1) intelligibility of each PB word list heard, 2) clarity of each list heard, and 3) attention demand to understand the words in each list heard. The results for the OH-58D were:

Pilots' ratings of INTELLIGIBILITY - ANR ON received significantly higher ratings ($n=11$, $R=0$, $p<<0.002$) than did the Stock SPH-4A helmet.

Pilots' ratings of CLARITY - ANR ON received significantly higher ratings ($n=12$, $R=6$, $p<0.01$) than did the Stock SPH-4A helmet.

Pilots' ratings of ATTENTION DEMAND - ANR ON received significantly better (less attention needed) ratings ($n=12$, $R=0$, $p<<0.002$) than did the Stock SPH-4A helmet. Note that a low value for attention demand is better than a high value.

A Chi-Square test was used to test for correlation between pilots' performance on the PB Word test and their speech ratings of intelligibility, clarity, and attention demand. For each of the three types of rating data, matched pairs for good listening level and for poor listening level were compared to the PB word data for the same matched pairs. The results of each comparison were sorted into three mutually exclusive categories for the Chi-Square test: 1) Rating data and PB Word data for good versus poor show a difference in the same direction, 2) there is no difference, 3) the difference is in the opposite direction. For Study 1 in the OH-58D Chi-Square values were 27.38, df=2 ($p<0.001$) for the comparison of PB Word data with intelligibility ratings; 46.5, df=2 ($p<0.001$) for the comparison of PB Word data with clarity ratings; and 30.88, df=2 ($p<0.001$) for the comparison of PB Word data with attention demand ratings. Thus the rating scales were validated by the PB Word performance data.

As can be seen in Figure 6, results of Study 2 in the Seahawk and of Study 3 in the EH-60 essentially replicated the findings from Study 1 in the OH-58D. The same statistical analyses were performed on these data as had been performed for Study 1. Figure 6 shows the PB Word data for these two studies. Again PB word intelligibility with ANR ON was significantly better than with the stock helmet ($n=7$, $R=0$, $p<0.05$), two-tailed test, for the Seahawk and similarly for the EH-60($n=14$, $R=8$, $p<0.01$), two-tailed test.

Figures 7, 8, and 9 also show the speech rating data for Studies 2 and 3. In the Seahawk, ANR ON resulted in higher intelligibility ratings compared to the stock helmet (n=7, R=0, p<0.05), higher clarity ratings (n=6, R=0, p<0.05), and ratings of less attention needed for ANR ON than for the stock helmet (n=8, R=0, p=0.01). And, as for the OH-58D, PB Word intelligibility was correlated with the speech ratings (Chi-Square=32.25, df=2, p<0.01). In the EH-60 these results were again replicated. ANR ON resulted in higher intelligibility ratings compared to the stock helmet (n=11, R=3, p<0.01), higher clarity ratings (n=10, R=0, p≤0.002), and ratings of less attention needed for ANR ON than for the stock helmet (n=14, R=3.5, p<0.002). PB Word intelligibility was correlated with the speech ratings (Chi-Square=30.49, df=2, p<0.01).

Given that each of three independent studies showed better intelligibility, higher speech ratings, lower attention demand and a significant correlation between PB Word Intelligibility and speech ratings, the data for all three studies were combined. The means for PB word intelligibility across all 12 aircrew are shown at the far right side of Figure 6. PB Word intelligibility was

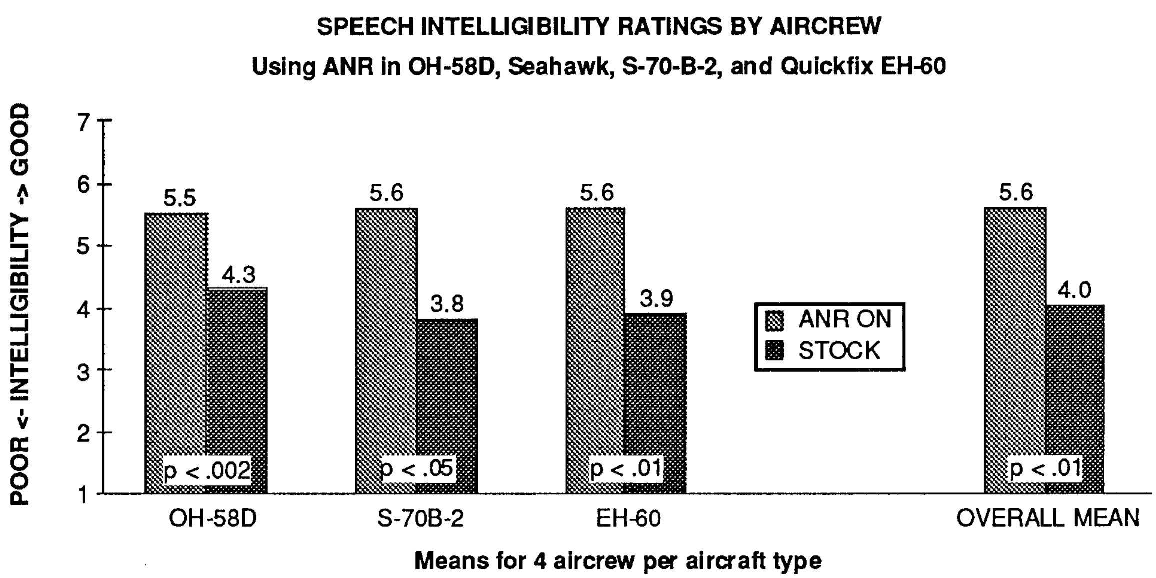


Figure 7: Intelligibility ratings for three aircraft for stock helmet compared to helmet with ANR ON.

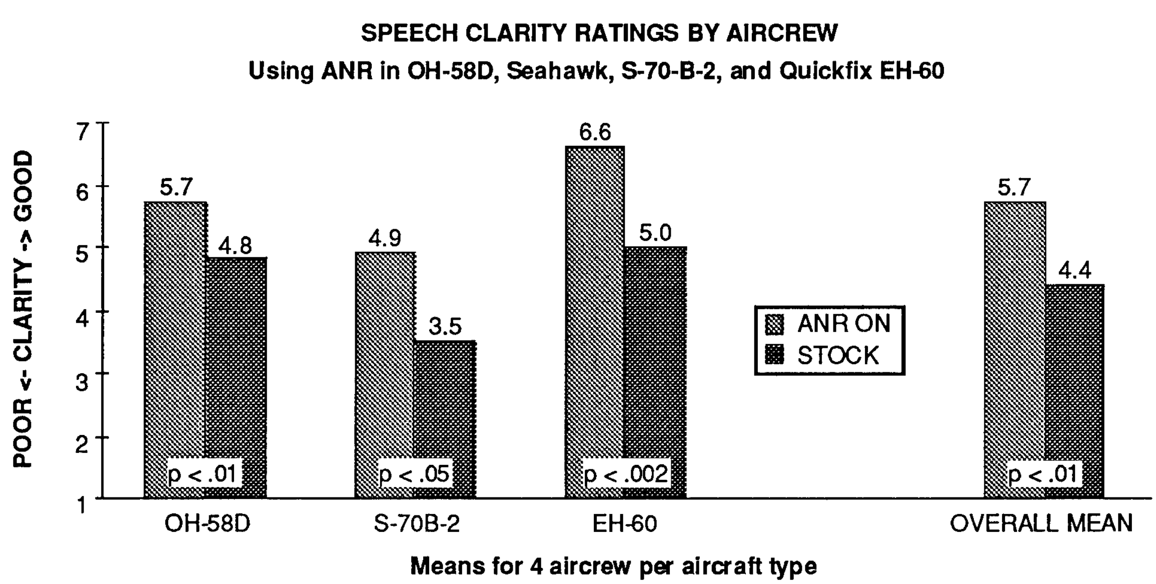


Figure 8: Clarity ratings for three aircraft for stock helmet compared to helmet with ANR ON.

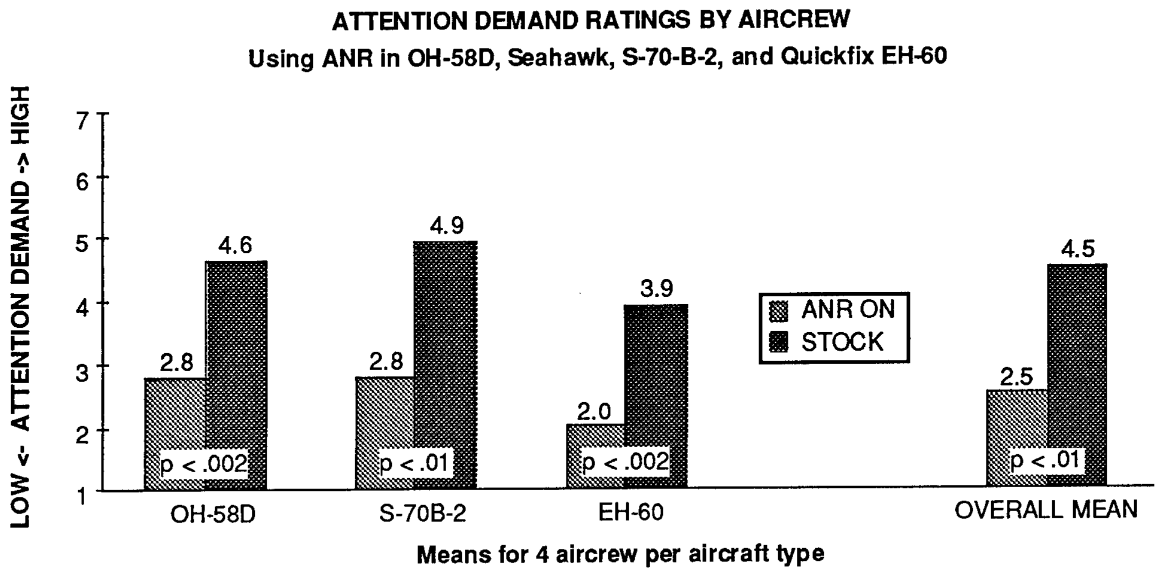


Figure 9: Attention Demand ratings for three aircraft for stock helmet compared to helmet with ANR ON. Note that low attention demand is desirable.

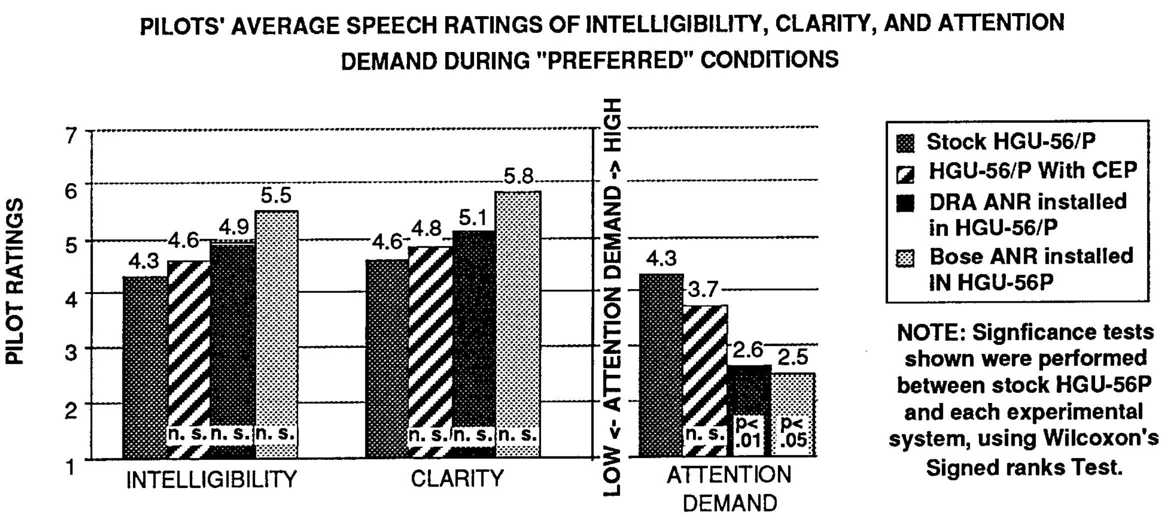


Figure 10: Ratings of Intelligibility, Clarity, and Attention Demand in NAH-1S Cobra for HGU-56/P stock, with CEP, with DRA ANR, and with Bose ANR.

significantly better with ANR ON (86.7%) than with pilots' stock helmets (71.4%) ($n=22$, $R=20.5$, $p<.01$, two-tailed test).

Figures 7, 8, and 9 show the corresponding overall means for intelligibility ratings, clarity ratings, and attention demand ratings, respectively. Again the ANR ON condition received significantly better ratings than the stock helmet condition for intelligibility ($n=18$, $R=4$, $p<0.01$), clarity ($n=18$, $R=1$, $p<0.01$), and attention demand ($n=21$, $R=3.5$, $p<0.01$).

Study 4 provided PB Word data, speech rating data, and operational suitability data from 6 Cobra Pilots flying the FLITE Cobra. Analysis of the digital audio tape recordings of the PB words as transmitted to the aircraft via radio revealed that, despite extensive efforts to ensure that the planned speech reception levels for each data run were achieved, the speech levels heard by the pilots varied by more than 3 dB from the intended levels. Thus the PB Word data were not usable.

Figure 10 shows speech rating data from Study 4 in the FLITE Cobra for each of four helmet configurations: Stock HGU-56/P, HGU-56/P with CEP, with DRA SPH-4B ANR, and with Bose HGU ANR, respectively. The means across six pilots are shown for each of the three rating scales: Intelligibility, Clarity, and Attention Demand. Wilcoxon's Signed Ranks Test for

Matched Pairs, two-tailed, was used to test for differences between each of the three experimental configuration (CEP, DRA ANR, Bose ANR) and the Stock helmet configuration. There were no significant differences among the intelligibility ratings nor among the clarity ratings, perhaps reflecting the variable speech levels during PB word radio transmission. There were, however, differences in attention demand.

The DRA ANR required significantly less attention compared to the Stock helmet (n=10, R=1.5, $p<0.01$). Similarly the Bose ANR required less attention than the Stock helmet (n=11, R=6.5, $p<0.05$). There was no significant difference in attention demand for the CEP compared to the Stock helmet (n=11, R=24, $p>0.05$).

3.4 Operational Suitability

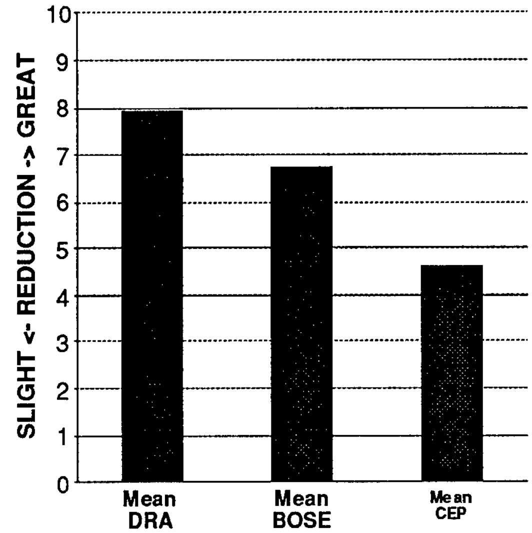
Data were collected on operational suitability, via the post-flight questionnaire, in each of the controlled studies of speech intelligibility (Studies 1-4). Results from Studies 1, 2 and 3 in the OH-58D, Seahawk and EH-60 compared DRA ANR to stock aircrew helmets and were essentially the same as the data that were obtained for that same comparison in Study 4 in the FLITE Cobra. Only the FLITE Cobra data are reported here dues to space limitations. The operational suitability data for Study 4 include not only the DRA ANR but also the Bose ANR and the CEP and are reported here.

Question 7 of the questionnaire asked pilots to rate the degree to which noise levels were reduced at the ear by the system they had just flown, as compared to their normal helmet. They indicated their responses by placing a mark at the appropriate position on a single horizontal line. Their responses were then mapped onto a 10 point scale and are shown in Figure 11. For the two ANR systems, all six pilots reported some reduction in noise levels. For the CEP, five of the six pilots reported some reduction and Pilot 1 stated there was no reduction with the CEP compared to his stock helmet. Averages of the pilots' ratings are shown in Figure 11. A rating of 10 indicates 'noise level greatly reduced' and a rating of 1 indicates 'noise level slightly reduced'. Average ratings, to the nearest integer, were 8 for DRA ANR, 7 for Bose ANR, and 4 for the CEP (see Figure 11).

Question 6 asked Pilots to rate the communication system just flown against their normal communication system. Results are shown in Figure 12. A rating of 10 indicates 'communication system improved' and a rating of 1 indicates 'communication system degraded'. Average ratings, to the nearest integer, were 8 for DRA ANR, 8 for Bose ANR, and 4 for the CEP (see Figure 12).

Question 17 asked Pilots 'Based on your flying experience, rate the utility of the 'system' for helping you achieve your missions, in comparison with your normal communications system'. A rating of 10 indicates 'utility improved' and a rating of 1 indicates 'utility degraded'. Average ratings, to the nearest integer, were 9 for DRA ANR, 9 for Bose ANR, and 4 for the CEP (see Figure 13).

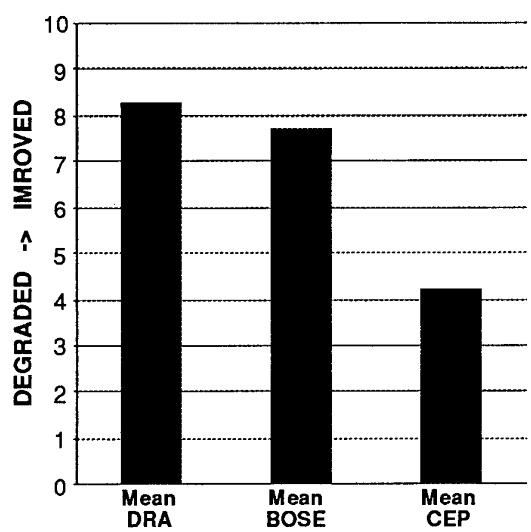
PILOT AVERAGE RATINGS OF HGU-56P WITH DRA AND BOSE ANR, AND CEP



Q7b: 'To what level did "system" reduce noise levels at the ear?'

Figure 11: Ratings of noise reduction for DRA ANR, Bose ANR, and CEP in HGU-56P helmet flown in NAH-1S Cobra. Differences among systems were not significant per Friedman's Test; k=3, n=6, X^2=.33.

PILOT AVERAGE RATINGS OF HGU-56P WITH DRA AND BOSE ANR, AND CEP



Q8a: 'How would you rate "system" vs. your normal comm system?'

Figure 12: Ratings of DRA ANR, Bose ANR, and CEP compared to normal ICS, as flown in NAH-1S Cobra. Differences among means were not significant per Friedman's Test; k=3, n=6, X^2=2.33.

Question 15 asked Pilots 'Do you think the system is acceptable for the operational environment?'. A rating of 10 indicates 'acceptable' and a rating of 1 indicates 'unacceptable'. Average ratings, to the nearest integer, were 9 for the DRA ANR, 9 for Bose ANR, and 3 for the CEP (see Figure 14).

3.5 Field Tests

In addition to the controlled studies of speech intelligibility, field tests of ANR were conducted in collaboration with operational Army aviation units. The speech rating scales, which had been previously validated against PB Word intelligibility, were used as well as the post-flight questionnaire to collect data on speech intelligibility with ANR and on operational suitability of ANR. These field studies were conducted during unit training missions on a non-interfering basis with training.

After flying training missions with ANR all but one pilot judged this ANR system to be ready for operational use. This pilot did not properly adjust his helmet to obtain an adequate seal of the earcups and had to turn off the ANR and use the standard communications. Therefore, he did not judge the ANR ready for operational use. However, in a follow-up flight, this pilot after properly adjusting his helmet, was able to use the ANR effectively and did rate the ANR ready for operational use, provided pilots were made aware of the need for adjusting their helmets.

In collaboration with A TRP, 4/6 CAV, Fort Hood, Texas, a test of ANR was conducted in the field during gunnery exercises, March 25-29, 1990. The objectives of this test were 1) to determine compatibility of the DRA Mk IV ANR system with the AH-64 IHADSS flight helmet, 2) to submit the ANR system to a wet, muddy environment like that of the Texas woods in the spring, 3) to obtain calibrated ratings of speech communications in the AH-64 during live fire training missions with the attendant high workload, 4) to determine any adverse effects on ANR performance of the weapons fire, specifically the 30 mm gun and 2.75" rockets, both of which produce acoustic noise peaks of short duration with rapid rise times and high levels, and 5) to determine the acoustic response of the ANR to weapons fire. The 30 mm gun has been reported by AH-64 pilots to be particularly disruptive to communications using their stock IHADSS helmets.

The two DRA Mark IV ANR systems were easily installed in the individual pilots' helmets in 20 minutes time by the A TRP Aircrew Life Support Equipment Technician. Both physical and electrical compatibility with the IHADSS helmet were thus established. Throughout the five days of gunnery exercises while A TRP was camped out in the field, during day and night operations in rain and generally high levels of humidity, both ANR systems functioned without failure. Participating pilots rated communications heard with ANR as more intelligible, clearer, and less demanding of their attention than the communications they normally

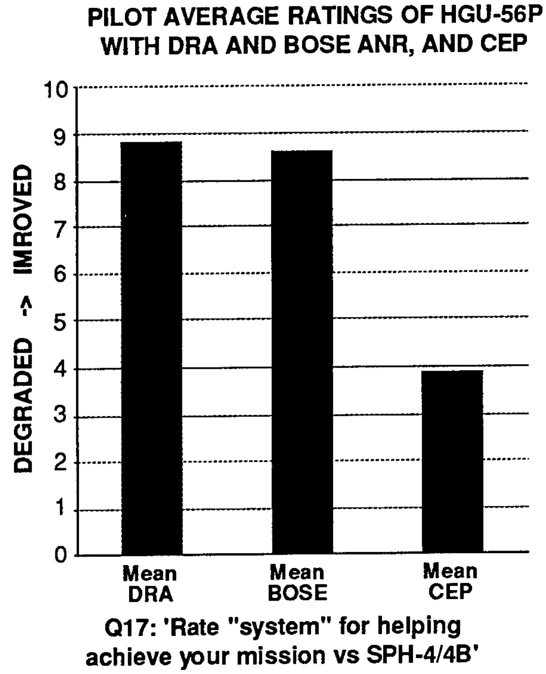


Figure 13: Ratings of DRA ANR, Bose ANR, and CEP for helping achieve mission success, as flown in NAH-1S Cobra Differences among systems were significant per Friedman's Test; $k=3$, $n=6$, $X^2=4.08$, $p<.05$.

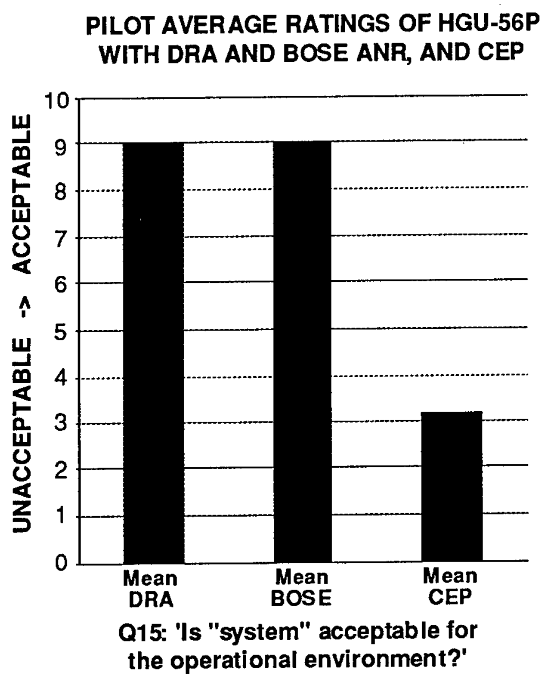


Figure 14: Ratings of acceptability for the operational environment for DRA ANR, Bose ANR, and CEP, as flown in NAH-1S Cobra. Differences among means were significant per Friedman's Test; $k=3$, $n=6$, $X^2=9.33$, $p<.01$.

experience with their stock IHADSS helmets. There were no adverse effects reported on ANR operation during live weapons firing.

OH-58D National Training Center Evaluation

An ANR helmet was taken to the simulated battlefield for operational testing by an OH-58D platoon under simulated combat conditions. Testing during this field study was conducted at the National Training Center (NTC) using a modified SPH-4A helmet fitted with ANR.

> NTC, at Ft. Irwin, California, is a thousand square mile section of the high Mojave Desert in a remote area northeast of Los Angeles. U.S. Army armored units rotate through NTC for combat training that is as close as it can be to actual war, fighting against an Opposing Force (OPFOR), relying on their equipment, training, and experience to maneuver as complete battalions and brigades, force-on-force. Lasers instead of bullets are fired during the battles, with computers keeping track of who has been "killed". The emotional stress of NTC for visiting units trying to beat the "home team" OPFOR is second only to that of actual combat [ref 12].

"Flights during this Phase were conducted at both Low-level, and NOE altitudes. The ANR equipped helmet provided clear, intelligible communications throughout all flight modes under simulated combat conditions." (pilot comment)

Desert Shield and Desert Storm Use

The most severe conditions under which test data were gathered occurred during both pre-combat and combat conditions involving Desert Shield and Desert Storm using the same SPH-4A helmet used at NTC. "During Desert Storm under actual tactical training conditions the ANR continued to show its usefulness by providing the aircrew members with clear intelligible communications, while reducing fatigue by reducing the attention necessary to understand incoming communications." (pilot comment)

"The SPH-4A helmet with ANR was worn with the M-24 Chemical Protective Mask with only one observed drawback: The mask obscured the crew member's view of the location of the ANR battery box / control switch. Training the crewmember on the location of the switch solved this problem. However, power source and switch location need to be looked into in greater detail. The straps of the M-24 did not interfere with the ANR system and did not cause the seal of the earcups to be broken." (pilot comment)

"The SPH-4A helmet with ANR had not been modified with a Thermal Plastic Liner (TPL). A decision was therefore made by the pilots to not fly the ANR during night combat missions for fear of developing "hot spots" while under Night Vision Goggle conditions and possibly jeopardizing a mission. " (pilot comment)

"The lack of a TPL in the ANR helmet resulted in it use during Desert Storm being limited to one flight during daylight conditions. During this mission, the ANR continued to operate as previously observed on earlier tests - providing increased intelligibility and reduced attention demand for voice communications. At no time during Desert Shield or Desert Storm did the system fail, despite the extreme temperatures and ubiquitous sand characteristic of the severe Persian Gulf area desert environment." (pilot comment)

The OH-58D Standardization Instructor Pilot had the following comments regarding the utility of ANR based on his experience in the Gulf War:

> The need for clear communications during air-combat operations was understood prior to the Gulf War, however the impact of good communications on mission success was realized time and again during both Desert Shield as well as Desert Storm when terrain and altitude took their toll on communications. The need to use low radio power settings while flying at very low altitudes (5 ft. to 25 ft.) at night with low or no illumination required aircrew to divide their attention between flying the aircraft and the incoming radio calls. Add to this situation, incoming radio calls on 3 to 5 radios, in some cases simultaneously, along with noises produced by multiple electric cooling motors within the cockpit, rotor blades, aircraft engine noises and it becomes obvious that the aircrew can become fatigued with this workload. The fact that ANR equipped helmets have proven to reduce the noise levels at the earcup and thereby lessen the necessary attention required to monitor incoming calls because of increased clarity allows aircrew to direct their attention to other aspects of the mission to insure complete success.
>
> *Mission*: Conduct Screen Operations to locate suspected enemy transmitter.
>
> *Friendly Situation:* Support supplied by 2/4th CAV, 24th ID (MECH)
>
> *Enemy Situation:* Suspected enemy transmitter located 5 Km inside friendly territory.
>
> *Synopsis:* During the conduct of this mission 2 OH-58 D aircraft teamed with a platoon of ground CAV M-2 Bradleys armored personnel carriers (APCs) and M-1 Abrams tanks were to search for and eliminate the suspected enemy transmitter. Communications during the linkup and execution of this mission was extensive due to the short planning time given to the mission. Overall the mission was successful, however the workload for the aircrew was exhaustive due to amount of communication traffic. Situations like this are where ANR could and would help the aircrew insure clear understanding of incoming radio traffic and therefore correct execution of instructions to avoid fratricide as was almost the case not once but 3 times during this mission alone. The high number of cases of fratricide during Desert Storm serves to show that all that could be done in insuring clear communications has not been accomplished. ANR can help in this area.

5 . DISCUSSION

Aircrew operating in modern rotary wing military aircraft such as the NAH-1S (Cobra), UH-1H (Huey), OH-58D (Kiowa), AH-64A (Apache), EH-60/S-70A-9 (Black Hawk) and the S-70B-2 (Seahawk) need to be provided with additional attenuation devices in order to:

(a) maintain reasonable manning levels for operational flying and meet current hearing conservation regulations which allow a Permissible Daily Noise Dose (PDED) of 85 dB(A) for an 8 hr day, and

(b) enhance communications (speech) intelligibility at the ear in order to improve mission task performance.

High ambient noise levels are generated in all rotary wing military aircraft. Ambient cabin noise levels in the S-70A-9 Black Hawk, for example, are in the order of 106 dB(C) to 109 dB(C) during cruise flight and are representative of those found in other aircraft. Aircrew helmets (such as the ALPHA and SPH-4 helmets) have generally good passive attenuation properties and serve to reduce the cabin noise level considerably before it reaches the ear. However, aircrew wearing helmets are exposed to at-ear SPLs higher than 85 dB(A). In the S-70A-9 during cruise flight, for example, a pilot wearing the ALPHA helmet still receives around 90 dB(A) at-ear, meaning he could only fly for 2 hr 31 min before exceeding his permissible daily exposure duration.

When taken in combination, the results of studies reported and reviewed in this paper suggest that:

(1) Both the DRA ANR and USAARL CEP systems effectively reduce at-ear SPLs, providing 10 and 9 dB of overall additional attenuation respectively.

(2) However, the two units do have differing spectral characteristics (i.e., provide different levels of attenuation at different frequencies). The ANR system provides better attenuation at lower frequencies (i.e., below 800 Hz) while the CEP system provides better attenuation at higher frequencies (i.e., above 800 Hz). Given that flight helmets already provide good attenuation at high frequencies (i.e, above 800 Hz), standard aircrew helmets (such as the ALPHA) fitted with the DRA ANR system would provide the most effective broadband attenuation. It should also be noted that it is less likely that the performance of the ANR system would be degraded in the field, due to the integrated nature of its installation. A growing body of evidence suggests that earplugs are rarely fitted 'properly' in the field and generally only provide some 35% of the attenuation reported under 'ideal' (properly fitted) measurement conditions [ref 3].

(3) In terms of speech intelligibility, present data show that in the field tests with three different aircraft, the SPH-4A and SPH-4B flight helmets do not meet MIL-STD-1472 for normally acceptable intelligibility under actual field acoustic conditions. In comparison, the SPH-4A and SPH-4B modified by the installation of the RAE Mark-IV ANR earcups, a 15-minute installation, was able to meet not only the normally acceptable intelligibility level but also approached the exceptionally high intelligibility level required in MIL-STD-1472 for operational systems.

(4) Ratings made by the pilots of speech intelligibility and speech clarity were highly correlated with measured PB word intelligibility, thus validating these rating scales as an independent test instrument. Additionally, pilots' ratings of attention demand to understand the speech were inversely correlated with PB word intelligibility, that is, the lower the intelligibility the more of the pilots' attention was needed to understand the speech, to the detriment of other flight tasks. The rating data indicated that the speech communications heard with the RAE Mark IV ANR earcups were more intelligible, clearer, and required less attention for understanding than those heard with the stock SPH-4A and SPH-4B helmet earcups. The high correlation of PB word intelligibility to the rating data allowed valid rating data to subsequently be collected from pilots while flying training combat missions.

(5) For the training missions flown with ANR, the ANR equipped flight helmets received ratings of higher intelligibility, higher clarity, and less attention demand than did the Stock SPH-4A helmets.

(6) When the DRA ANR and Bose ANR systems were compared to the triple flange version of the CEP, it was found that pilots judged that all three systems reduced noise levels at the ear compared to stock flight helmets. Also, all three systems received ratings from pilots which indicated that the speech received was clearer and more intelligible than the speech heard with stock helmets. There were, however, differences in the ANR systems compared to the CEP in two important domains: 1) operational suitability and 2) expected interaction of each system's attenuation frequency response with speech perception and masking of audio signals.

(7) When pilots were provided equal flight time and equal flight maneuvers for CEP and for ANR, their judgements of operational suitability for ANR were significantly higher than for CEP. In fact, the CEP was rated on the 'unacceptable' side of the operational suitability scale. The potential problems that pilots noted for CEP were a) fragility, b) discomfort, c) excessive helmet donning time, d) high replacement costs due to expected frequent wire breakage, e) ear irritations and infections in the harsh combat environment, f) potential for snagging the thin wires and pulling out the CEP, and g) variability in ICS speech levels due to the CEP shifting inside the ear canal during flight.

In contrast, the only operational problems noted by pilots for ANR were a) need to supply power via ships power so as not to be dependent on batteries, b) need to teach pilots to fit their helmets well so as to prevent the ANR earcups from breaking seal. Both of these problems can be easily addressed. Once the pilots in the studies reported here had learned to adjust their helmets, they had no problems with breaking seal on the ANR.

Sources of power for ANR will have to be determined for each helicopter in the fleet. Some, like the OH-58D, already have a clean 28V supply for night vision goggles. Others will require an adapter to condition the voltage source. However, such a modification is relatively minor in comparison to, say, installation of a new radio or new ICS.

(8) A side benefit of pilots having to properly adjust their helmets in order to use ANR is that the passive attenuation of the helmets would be optimised.

(9) The other difference between ANR and CEP which impacts on speech intelligibility is the frequency spectrum of their respective attenuation curves. The better low frequency attenuation of ANR should produce a greater improvement in speech intelligibility because of ANR's greater reduction in the upward masking effects of low frequency noise [ref 6].

6 . SUMMARY

Combined work by US AFDD, AS AMRL, and Human Research Engineering Division (HRED) of the U.S. Army Research Laboratory (ARL) and has shown that ANR has proven itself to useful in reducing crewmember work load by effectively reducing at-ear sound pressure levels by around 10 dB, improving clarity and intelligibility and reducing attention demand within the crewstation during both peace time tactical training evaluations as well as combat situations. These factors are key elements in mission success. The DRA ANR and the Bose ANR were both rated as highly acceptable for operational use by US Army and Australian aircrew. While the CEP also reduces noise and improves speech intelligibility, certain of its design features are inherently not as suitable for the operational environment, and its attenuation characteristics are expected to be less enabling than ANR for accurate detection and perception of speech and other audio signals in the cockpit. For these reasons, ANR is the preferred technology for improving cockpit communications and reducing cockpit noise levels for military aircrew.

REFERENCES

1. Australian Standard 1269-1989 (1989). 'Acoustics — Hearing Conservation', Standards Association of Australia, Sydney, Australia.

2. ANSI, (1960). 'Method for Measurement of Monosyllabic Word Intelligibility', USA Standard, Report No. ANSI S3.2-1960 (R1971), US American National Standards Institute: 1430 Broadway, New York, NY 10018, 1960.

3. Berger, E.H., Franks, J.R. and Lindgren, F. (1996). 'International Review of Field Studies of Hearing Protector Attenuation', Chapter 29 in Axelsson, Borchgrevink, Hammernink, Hellstroem, Henderson and Salvi (Eds), Scientific Basis of Noise Induced Hearing Loss, Thieme, New York.

4. Campbell, R.A. (1965). 'Discrimination Test Word Difficulty', Journal of Speech and Hearing Research, 8, pp. 13-22.

5. Cooper, G. E., and Harper, R. P. (1969). 'The use of pilot rating in the evaluation of aircraft handling qualities', NASA Technical Note TN-D-5153, National Aeronautics and Space Administration, Ames Research Center, Moffett Field, California, USA.

6. Crabtree, R.B. (1996). 'Constraints in the Application of Personal Active Noise Reduction Systems', Conference PrePrint, Audio Effectiveness in Aviation, sponsored by Aerospace Medical Panel, AGARD, NATO, Copenhagen, Denmark, 7-10 October, 1996.

7. Egan, J.P. (1948). 'Articulation Testing Methods', Laryngoscope, 1948, 58, pp. 995-91.

8. Fairbanks, G. (1958). 'Test of Phonemic Differentiation: The Rhyme Test', Journal of the Acoustical Society of America, 1958, 30, pp. 596-600.

9. Garinther, G. and Peters, L. 1990. 'Impact of Communications on Armor Crew Performance', Army Research and Development Acquisition Bulletin, January-February, 1990, pp. 1-4.

10. Gossett, T.D., Simpson, C.A. and Gardner, D.G. (1992). 'Active Noise Reduction (ANR) Field Test in EH-60 Quickfix Helicopters', Report for Program Manager-Advanced Quickfix, Project Manager Signals Warfare, 9 November, 1992.

11. Guilford, J.P. (1936). 'Rating Scale Methods', Psychometric Methods, Chapter IX, pp. 263-284, McGraw Hill, New York and London.

12. Halberstadt, H. (1989). 'NTC: A Primer of Modern Land Combat', Presidio Press: Novato, CA, USA, 1989.

13. House, A.S., Williams, C.E., Hecker, M.H.L and Kryter, K.D. (1965). 'Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set', Journal of the Acoustical Society of America, 37, pp.158-66.

14. International Electrotechnical Commission (1981) Standard 711. 'Occluded Ear Simulator for the Measurement of Earphones Coupled to the Ear by Ear Inserts', International Electrotechnical Commission (1981), Geneva, Switzerland.

15. King, R.B., Saliba, A.J., Creed, D.C. and Brock, J.R. (1996). 'Assessment of Noise Levels in and Around the Sikorksy S-70A-9 Black Hawk Helicopter', Defence Science and Technology Organisation Technical Report 300, DSTO-TR-0300. Aeronautical and Maritime Research Laboratory, Melbourne, Australia.

16. King, R.B., Saliba, A.J., Brock, J.R. and James, S.H. (1996). 'Assessment of a Helmet Mounted Active Noise Reduction System in the S-70A-9 Black Hawk Helicopter', Defence Science and Technology Organisation Technical Report TBA, DSTO-TR-TBA. Aeronautical and Maritime Research Laboratory, Melbourne, Australia (in press).

17. King, R.B. and Foran, D.A. (1992). 'In-flight evaluation of noise levels and assessment of active noise reduction systems in the Seahawk S-70B-2 helicopter', Aeronautical Research Laboratory Technical Report 9, ARL-TR-9. Aeronautical and Maritime Research Laboratory, Melbourne, Australia.

18. Lehiste, I. and Peterson, G. (1959). 'Linguistic Considerations in the study of Speech Intelligibility', Journal of the Acoustical Society of America, 1959, 31, pp. 280-86.

19. Mozo, B.T. and Ribera, J.E. (1996). 'The Communications Earplug: A Logical Choice for Voice Communications in Aircraft', Conference Preprint, Audio Effectiveness in Aviation, sponsored by Aerospace Medical Panel, AGARD, NATO, Copenhagen, Denmark, 7-10 October, 1996.

20. Schooneveldt, G., King, R.B. and Craven, D. (1995). 'Recording Aircraft Acoustical Environments for Auditory Display Design', in Oldfield, S. (Ed.) 'Proceedings of the TTCP Workshop on the Integration of Sound with Visual Displays' (TTCP/SGU/96/tba). Washington, DC, USA: The Technical Cooperation Program, Subgroup U, Technical Panel 7.

21. Simpson, C.A. and Gardner, D.R. (1991). 'Operational Speech Intelligibility Using Active Noise Reduction in Army Helicopters During Training and Actual Missions', Paper Presented at the Avionics Technical Symposium, American Defense Preparedness Association, Eglin Air Force Base, Florida, November 20-22, 1991.

22. Simpson, C.A. and Gardner, D.R. (1994). 'Flight Test of active Noise Reduction (ANR) in Army Combat Helicopters', Briefing given to Tri-service Aircrew Life Support Meeting, Joint Active Noise Reduction (ANR), Passive Noise Reduction (PNR) Working Group, Brooks AFB, 31 August, 1994.

23. Simpson, C.A. and Williams, D.H. (1992). 'Operational Speech Intelligibility Using Active Noise Reduction in Royal Australian Navy Seahawk Helicopter', Working Paper for Crew Station R&D Branch, Simulation and Aircraft Systems Division, Aeroflightdynamics Directorate, US Army AVSCOM, 15 May, 1992.

24. US Army, (1987). 'Aeromedical Training for Flight Personnel', FM 1-301, Headquarters, Department of Army, May 1987.

25. US Department of Defense, (1991). 'Department of Defense Instruction 6055.12', Hearing Conservation Program, 26 March 1991.

26. Voiers, W.D. (1977). 'Diagnostic Evaluation of Speech Intelligibility', Chapter 34 in Hawley, M. (Ed.), Speech Intelligibility and Speaker Recognition, Dowden, Hutchinson & Ross: Stroudsberg, PA, USA, 1977.

27. Whitaker, L.A. (1994). 'Speech Intelligibility and Task Performance: Phase II SBIR Interim Report', prepared under contract to the US Army Research Laboratory, Aberdeen Proving Ground, MD, 26 August 1994.

NEXT GENERATION ACTIVE NOISE REDUCTION SYSTEMS

P Darlington[1] & G M Rood[2]

(1) Department of Acoustics and Audio Engineering
University of Salford
Salford, M5 4WT, UK

(2) MMI Dept
DRA, Room 204, Bldg R177
Farnborough, Hants GU14 6TD, UK

Active techniques for attenuating the sound pressure levels at the ears of aircrew are examined. Conventional Active Noise Reduction (ANR) systems are reviewed. Their performance is shown to be constrained by their essential "feedback" architecture. ANR systems which avoid the feedback path are introduced and the performance of a new active noise reduction system is reported. The new system is demonstrated to offer such attenuation of noise that hearing damage risk is significantly reduced and operational performance enhanced.

1 INTRODUCTION

Occupants of military vehicles can be subjected to high levels of noise. This noise degrades the efficiency of speech communication by masking, acts as a stressor, and threatens audiological health. Although personnel may wear protective headgear and earmuffs, which offer some protection, noise levels at the ear remain a significant problem. Active techniques to reduce the noise levels at the ear offer an attractive alternative to attempting to increase the passive attenuation afforded by a helmet / headset combination, particularly at low frequencies.

Active hearing defenders based upon analog electronics have been reported over the last 20 years, but their performance is limited by stability and complexity constraints. The peak attenuation offered by such systems in practical realisations in circumaural muffs is around 20 dB. The operating bandwidth of these devices is lowpass limited to frequencies below 1 kHz..

Research and development at DRA Farnborough has focussed upon the goals of increasing attenuation and extending bandwidth of practical active noise reduction (ANR) systems. This work has generated a system which uses digital techniques to achieve significantly better performance than contemporary analog systems. It is the purpose of this paper to review the conventional analog ANR system to identify those factors which limit its performance, to introduce the new DRA digital noise reduction system and to outline areas of research which are informing the development of other next generation ANR systems.

2 BACKGROUND

The conventional ANR system (Figure 1) detects the pressure close to the ear using a miniature microphone. The pressure, p_m, is associated with noise generated by the communications telephone, p_t, and the unwanted noise caused by transmission of cabin noise, p_n, such that the microphone voltage, v_m, is:

$$v_m = M(p_t + p_n) \qquad (1)$$

where M is the microphone transfer characteristic (V/Pa). The sound generated by the telephone has two components; that due to the drive voltage from the communication system, v_s, and that due to the drive voltage from the ANR system, v_c. Assuming a telephone characteristic of T (Pa/V), the pressure induced by the telephone at the position of the microphone is:

$$p_t = T(v_c + v_s) \qquad (2)$$

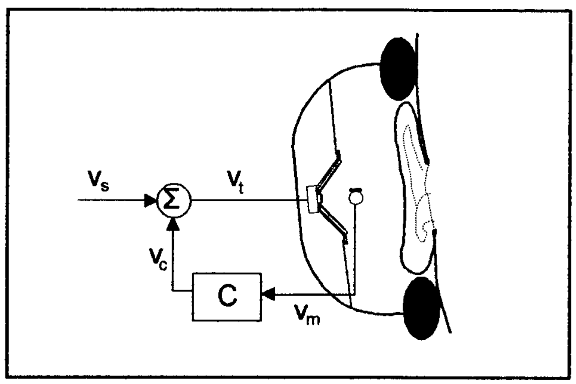


Figure 1 The conventional ANR system

The control system voltage of a conventional ANR system, v_c, is derived by operating upon the voltage detected by the sense microphone by a filter C:

$$v_c = C.v_m \qquad (3)$$

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

This gives the ANR system the structure of a canonical "feedback" controller.

Combining equations (1) - (3) gives an expression for the pressure at the microphone location:

$$p_m \left[= \frac{v_m}{M} \right] = \frac{(v_s T + p_n)}{(1 - MCT)} \qquad \textbf{(4)}$$

Equation 4 states that the pressure at the sense microphone in the absence of the controller (v_sT+p_n) is scaled by the frequency domain factor $(1\text{-}MCT)^{-1}$ when the controller is turned on. For the controller to attenuate the sound, the "compensating" filter, C, must be designed with respect to the fixed electroacoustics, MT, such that the magnitude of $(1\text{-}MCT)^{-1}$ is smaller than unity. The attenuation produced is:

$$att.\ (dB) = -10\log\left(\left|\frac{1}{(1-MCT)}\right|^2\right) \qquad \textbf{(5)}$$

Note that the ANR system operates upon both the unwanted noise and the pressure generated by the communication system by the same amount - for this reason *a conventional ANR system does not improve the signal to noise ratio* (although speech intelligibility may be significantly improved by the change in the noise spectrum once masking is accounted for). Practical systems pre-emphasise the speech signal by filtering v_s by an approximation of 1-MCT.

The available attenuation shown in equation 5 is maximised when the open loop gain (magnitude of MCT) is maximised. However, the attenuation is practically limited by the requirement for the system to be stable, which demands that the open loop gain is smaller than unity at certain frequencies (when the phase of MCT is an integer multiple of 360°). There is seen to be a conflict in the specification of C; high gain required for high attenuation, low gain required for stability. It is this conflict which practically limits the performance of conventional ANR systems.

A further factor frustrates the performance of conventional ANR systems; noise enhancement. If the open loop traverses more than 360 degrees of phase (which is inevitable given the complexities of the electroacoustics MT) then, for any stable system,

there are frequencies at which the system enhances the pressure at the microphone position.

These regions of negative attenuation are unavoidable in conventional ANR systems. Despite these limitations, analogue ANR provides significant benefits where low frequency passive attenuation is difficult to achieve in practice using passive techniques. Many laboratory experiments have been reported and peak attenuations in the region of 20 dB are readily achievable. The high frequency noise enhancement can be reduced (at the expense of loosing low frequency attenuation) by reducing the system loop gain; a compromise between low frequency attenuation and high frequency enhancement has to be made.

Typical performance of a conventional ANR system is reported as Figure 2, which shows practically attainable active attenuation averaged over flight trials by RAE Farnborough, in a Sea Harrier fighter and strike aircraft and Sea King helicopters.

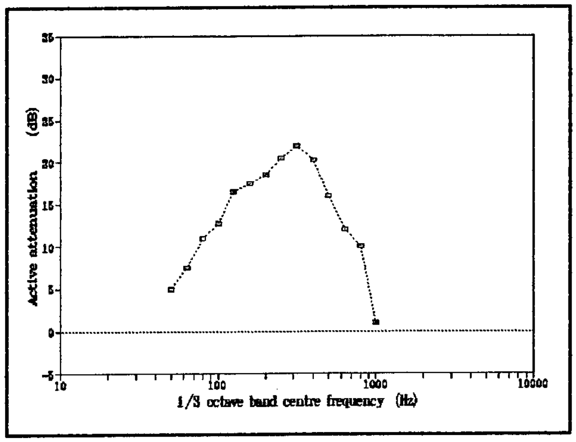


Figure 2 Typical performance of a conventional ANR system (average of trials in Sea Harrier and Sea King aircraft types).

The performance of a standard ANR system fitted to different aircrew's helmets will vary from wearer to wearer as a result of the slightly modified electroacoustics. An indication of the range of active attenuation provided in an operational context is reported as Figure 3, which shows minimum and maximum attenuations measured on 10 aircrew in Sea Harrier operational squadron trials.

The stability and enhancement problems which constrain the performance of conventional ANR systems stem from their inherent feedback structure. An obvious route for the development of improved ANR systems is to attempt to remove the feedback path.

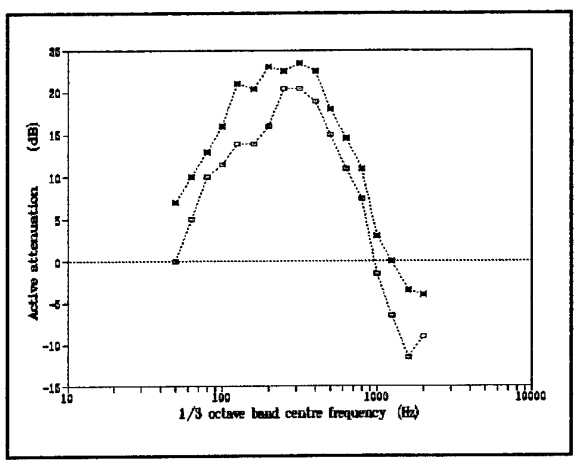


Figure 3 Maximum and minimum active attenuations measured on 10 Sea King aircrew in an operational squadron trial of the conventional ANR system

3 FEEDFORWARD ANR SYSTEMS

The feedback was introduced to the ANR system described in section 2 by including the total sense microphone voltage as a factor of the control voltage equation (3). Since the microphone is responsive to the pressure generated by the telephone, the feedback loop is generated. Two strategies for removing this feedback path in the context of an ANR system are considered below.

3.1 ANR with external pressure reference.

If a second microphone is placed outside the earmuff, it will respond to the pressure, p_{ext}, which is causing the noise component of the pressure inside the muff, p_n, but will be (relatively) insensitive to the output of the telephone as a result of the passive attenuation of the muff. If the pressures are related by a transfer function H:

$$p_{ext} = H^{-1}p_n \tag{6}$$

then the control voltage can be designed by feedforward operations through compensator W on the external pressure signal:

$$v_c = W.M.p_{ext} \tag{7}$$

The optimal configuration of W is:

$$W_{opt} = \frac{-H}{MT} \tag{8}$$

Unfortunately, the sound field in many vehicles is partially diffuse, such that the pressures at a single point outside a muff and at the sense microphone inside the muff are not perfectly correlated [1]. Under this situation, the optimum attenuation produced by the feedforward ANR system with external reference is limited to:

$$att.\ (dB) = -10.\log(1 - \gamma^2) \tag{9}$$

in which γ^2 is the ordinary coherence between the pressures at the external and internal microphones.

3.2 The Feedback Cancelling ANR System

An alternative approach to the removal of the feedback path in the ANR system of system 2 is to explicitly cancel the component of the sense microphone response due to the telephone electronically (the telephone drive voltage being known to the control system). The resulting control voltage has form:

$$v_c = W.(v_m - (MT)'.v_t) \tag{10}$$

in which W is the feedforward compensator, and MT' is filter representing an approximation of the electroacoustics MT. This results in the system depicted as Figure 4, which is an example of a control system structure known as *internal model control* ("IMC") [2,3].

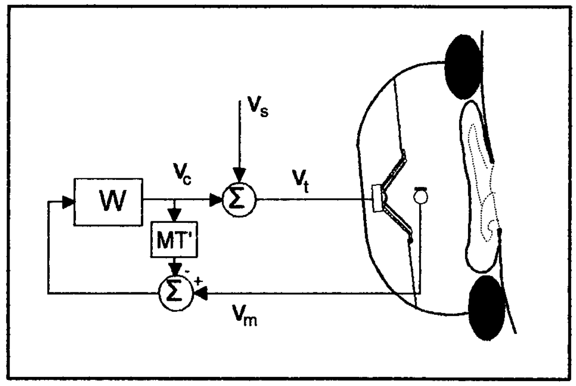


Figure 4 The feedback cancelling ANR system

In the ideal case where MT' is an exact match of the true electroacoustics, MT, then the control voltage is a function only of the noise component of the pressure at the microphone:

$$v_c = W.v_n \tag{11}$$

in which case the system has pure feedforward structure - the feedback loop has been removed by cancellation. The pressure at the sense microphone is then given by:

$$p_m = T.v_s + p_n.(T.W.M + 1) \tag{12}$$

The communication component of the signal generated by the telephone is not influenced by the ideal feedback cancelling ANR system and the noise, p_n, is scaled by (WTM+1). The optimal configuration of the controller W would be:

$$W_{opt} = \frac{-1}{MT} \tag{13}$$

and such a controller would theoretically yield infinite attenuation. Unfortunately, the optimal solution defined in equation 13 cannot be physically implemented, as the telephone response (pressure at the sense microphone divided by voltage at the telephone) includes a delay and so cannot be causally inverted. The causally constrained inverse of MT differs from the true inverse in such a way that the actual attenuation will be a function of MT and the statistics of p_n (see section 4).

In the practical case where MT' is not a perfect match of MT, an error transfer function can be defined:

$$E \doteq MT - MT' \tag{14}$$

in which case the pressure at the sense microphone is:

$$p_m = \frac{Tv_s + p_n(T.W.M + 1 - \frac{W.E}{T})}{(1 - \frac{W.E}{T})} \tag{15}$$

The communication system component of the signal generated by the telephone is scaled by $(1\text{-}WE/T)^{-1}$ and the noise is attenuated by:

$$att.\ (dB) = -10.\log\left|\left|\frac{TWM+1-\frac{WE}{T}}{1-\frac{WE}{T}}\right|^2\right| \quad \textbf{(16)}$$

As the error in the feedback cancellation path, E, is reduced, the disruption of the communication signal, p_s, reduces, and the attenuation of p_n approaches the attenuation associated with a causally constrained estimate of -1/MT. As MT represents a physical electroacoustic path, it can be modelled with arbitrary accuracy, such that E can be made as small as available technology allows.

4 EXPERIMENTAL VALIDATION

A new ANR system has been constructed, using the feedback cancelling (IMC) structure described in section 3, to avoid the problems of feedback systems identified in section 2. The new system has a mixed analog / digital controller, as illustrated in Figure 5. The inner loop is entirely analog and is itself an ANR system. This inner loop serves to reduce the length of the impulse response of the transfer function between the digital control input to the telephone, $v_{c,d}$, and the response from the microphone, v_m. This impulse response is approximated by a digital finite impulse response ("FIR") filter to cancel the feedback loop and the computational cost of implementing this filter reduces as the impulse response of the closed analog inner loop reduces in length.

The feedforward control filter W is adjusted using adaptive methods. A Widrow-Hoff LMS algorithm adjusts the weights of the FIR filter W, with filtered-x compensation for the dynamics of the plant. The system has been found to be robustly stable, in contrast to other reported IMC based ANR systems built around open ear headsets.

A number of laboratory trials were conducted at Farnborough in 1993 on the new digital ANR system in broad band Harrier and Tornado cabin noise. With the adaption parameters adjusted for each of 7 subjects, average active attenuations peaked at 33 dB, giving an indication of the system's performance. Later trials were carried out in September 1994, using fixed adaption parameters, again in a broad band noise spectrum representing Harrier/AV8B cabin noise. The results are reported as Figure 6, with average attenuations peaking at 34 dB, achievable in the 400 and 500 Hz bands. Comparable enhancement to the analogue feedback system is present in the octave above 1kHz., but the variance of the measurements is usefully in line with passive attenuation data and that from conventional ANR systems.

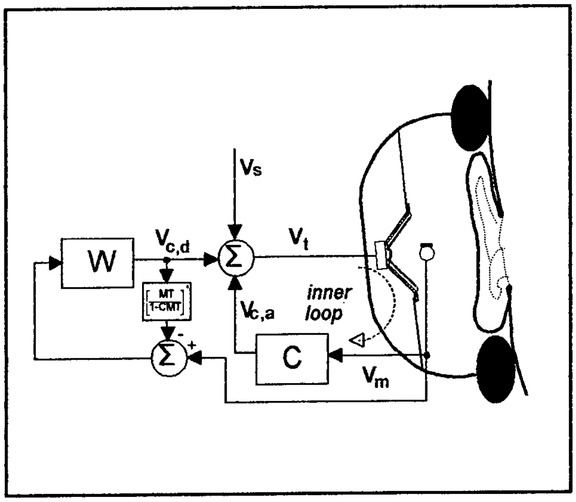


Figure 5 The new DRA Farnborough active noise reduction system

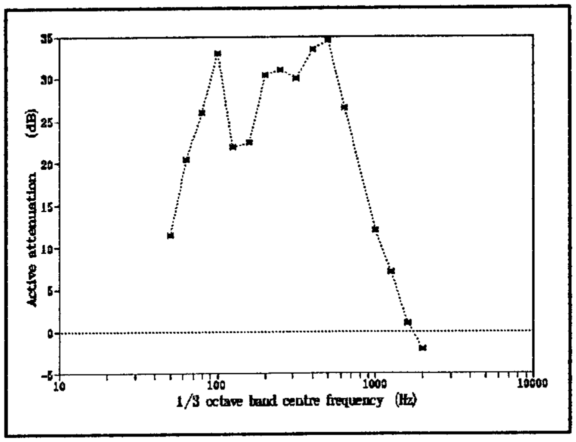


Figure 6 Performance of the new ANR system (mean of ten subjects, measured in Harrier AV8/B noise)

5 FURTHER DEVELOPMENTS

Despite the performance advantages achieved in the new system, the authors are pursuing further developments in active noise reduction in the context of aircrew helmets. Current work is focussed upon extending the bandwidth of operation of the system and optimising ANR systems with respect to psychoacoustic criteria. The bandwidth extension studies are currently addressing modifications of the electroacoustics and imperfections in the practical operation of the IMC feedback cancelling filter. The optimisation work is using Genetic Algorithms to suggest novel configurations of ANR systems. These configurations may attempt, for example, to maximise A weighted active attenuation over a specified bandwidth in a specified noise field. This allows the compromise between high frequency enhancement and low frequency active attenuation, described in section 2, to be made automatically in a manner which gives maximum benefit to the user.

6 CONCLUSIONS

Although conventional analog ANR systems provide useful levels of active attenuation to combat the low frequency noise present under the flying helmets of aircrew, contemporary practical realizations of such systems are close to their optimum achievable performance. Measurements during operational flight trials have shown that some risk of hearing damage remains in a number of aircraft, even with conventional ANR systems incorporated. These measurements also show that, in noisy aircraft, improvements in active attenuation can further improve speech and signal intelligibility. The active attenuations achieved by the new digital ANR system described in this paper are sufficient to essentially remove the risk of hearing damage [4,5], and to significantly enhance operational performance - particularly in crew communications.

7 REFERENCES

[1] P Darlington, P D Wheeler & G A Powell "Adaptive noise reduction in aircraft communication systems", Proc. ICASSP '85, 716-719, 1985

[2] S J Elliott & T J Sutton "Feedforward and feedback methods for active control", Proc. IOA **16**(2), 255-276, 1994

[3] B Rafaely & S J Elliott "Adaptive internal model controller -stability analysis", Proc. Internoise '96, 983-988, 1996

[4] Susan H James & G M Rood "In-flight assessment of a helmet-mounted active-noise reduction system in Sea King 5 helicopters" (Unpublished Technical Memorandum)

[5] S H James & K M Harpur "In-flight assessment of a helmet mounted active noise reduction system in Sea Harrier FRS1" (Unpublished Technical Memorandum)

Aircraft Noise Profiles and the Efficiency of Noise Protection Devices in the Royal Danish Air Force

S Vesterhauge*, PA Osterhammel*, A Nørby Rasmussen*,
JNS Oldenburg☼ and ES Jensen☼.
*Dept. of Otolaryngology, Rigshospitalet #2074, DK-2100 Copenhagen, Denmark;
☼The Royal Danish Air Force, Sqdr. 590, Vaerloese AFB; DK-3500 Vaerloese, Denmark.

INTRODUCTION:

Except for being hazardous to the function of the ear itself, noise has a lot of unpleasant non-organic capabilities. It is annoying, noise interferes with performance and efficiency, and it interferes with communication. No matter what we do, we all have to live with and accept certain levels of noise. This, indeed, counts for aviation too. It has been told, that when Louis Blériot in 1909 flew from France to England, the noise from his 25 HP engine heard from the ground by those fortunate enough to witness this historic event, was probably 20 to 30 dB louder than the noise reaching the ground from a current jet aircraft[1]. This was caused by the fact that Blériot flew very much lower than modern aircraft. So, due to simple physical laws, the closer you are to a noise source, the more you are exposed. And those closest to an aircraft are those working in it or outside the plane. In the air force and in other flying units of our defence, personnel is exposed to high levels of noise.

The purpose of the present study, is simply to map, in a comparable way the noise impact on personnel working at different positions in relation to aircraft used by the Danish defence - to establish the efficiency of different noise protection devices used by personnel working at different positions - and finally to advice the proper authorities concerning the proper use of noise protection devices in order to avoid as much as possible the harmful effects of aircraft noise as described above.

METHODS:

For the noise measurement we used a digital audio tape recorder made by Sony and a Brüel & Kjær sound-level-meter. The tape recorded noise signals were stored for later use and analysis in our laboratory. The analysis was performed on a Brüel & Kjær Audio Analyser providing information about the A-weighted noise level and the linear frequency spectra of the noise recorded. The measurements were made according to ISO 5129[2].

Until now, noise has been recorded and analyzed from the following aircraft:

T-17: SAAB Supporter. A one-engine, two-seated propeller aircraft used for training and reconnaissance,

C-130: Lockheed Hercules, know by everyone,

G-III: Gulfstream SMA-3. An American corporate transport aircraft, modified for fishery patrol operations

S-61: Sikorsky helicopter, mainly used for search and rescue operations, and

the *Westland Lynx* helicopter, used for maritime patrol and SAR operations.

In all cases, evaluation of the noise attenuating properties of helmets and head-sets was performed in a sound proof room in our laboratory, presenting on the first hand what we decided to be a standard aircraft noise, the Hercules noise, through loudspeakers surrounding the test person. A small silicone tube was inserted into the ear canal and the tip was placed between the noise protection device and the ear drum. In this set-up, the tube is connected to an external microphone and the difference between the sound pressure in the ear canal and the external sound pressure is recorded. The measuring method and equipment used is exactly the same as that used for insertion gain measurements in patients when fitted with hearing instruments. In the present case the gain measured as the result of the use of a noise protection device is negative and not positive as expected when a hearing instrument is fitted.

RESULTS:

Aircraft Noise:

In the case of the *T-17*, measurements were performed at the wing tip during motor tests. The A-weighted sound pressure was as high as 105 dB (fig. 1).

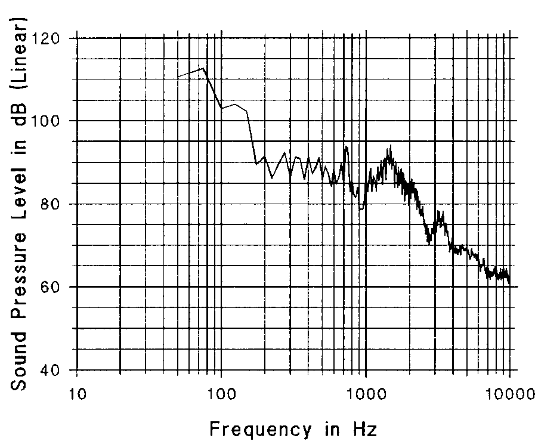


Figure 1. T-17 noise spectrum, measured at wing tip.

When measured at the level of the pilots ear in the cockpit during cruise at 1000', the A-weighted noise level is 94 dB and the frequency distribution is now characteristic of a propeller-driven aircraft, dominated by low frequency noise (see fig.2, next page).

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

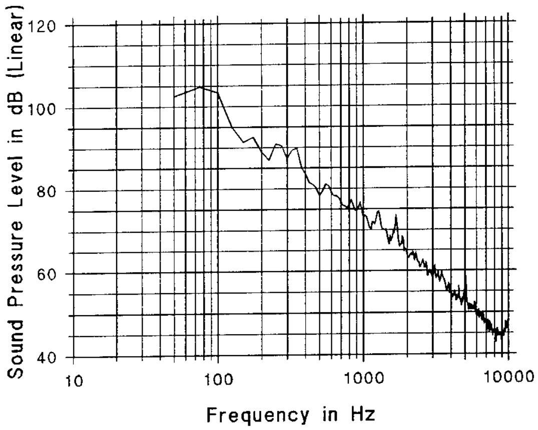


Figure 2. T-17. Noise spectrum measured during cruise.

In the *Hercules*, the noise level at the pilot seat is 91 dB(A) and the spectral pattern is still characteristic of that of a propeller driven aircraft (see fig 3, below).

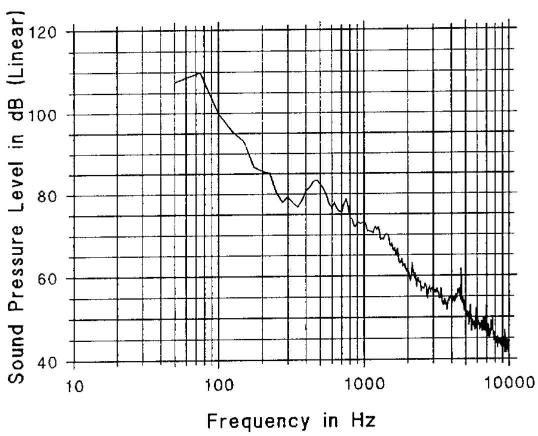


Figure 3. C-130. Noise spectrum in the pilot's seat measured during cruise.

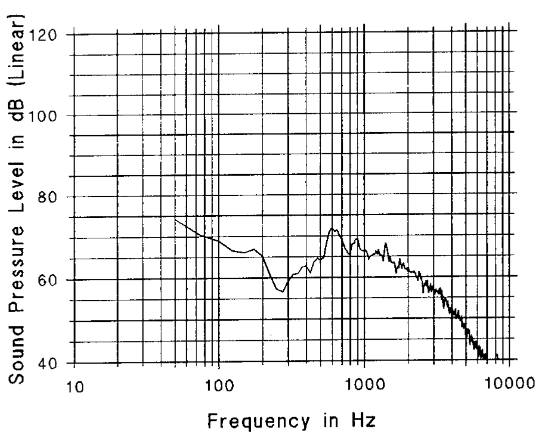


Figure 4. G-III. Noise spectrum at the flightdeck.

In the *Gulfstream-3, at flightdeck* (fig. 4), the noise level is the lowest measured level during this investigation. A 83 dB(A) level is considered safe for more than eight hours daily exposure. The relatively low noise level is caused by the fact that the turbofan engines are situated as far as possible from flight deck - in the rear end of the aircraft.

Moving in the aft direction of the cabin means approaching the power plants and considerably increasing the noise level (see fig. 5).

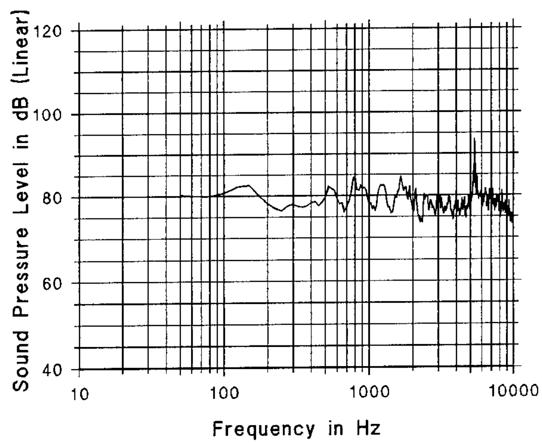


Figure 5. G-III. Noise level in the cabin.

In the *Sikorsky helicopter*, the sound pressure level is almost as high as that measured at the wing tips of the T-17, 104 dB(A), dominated by low frequency sound energy (fig. 6).

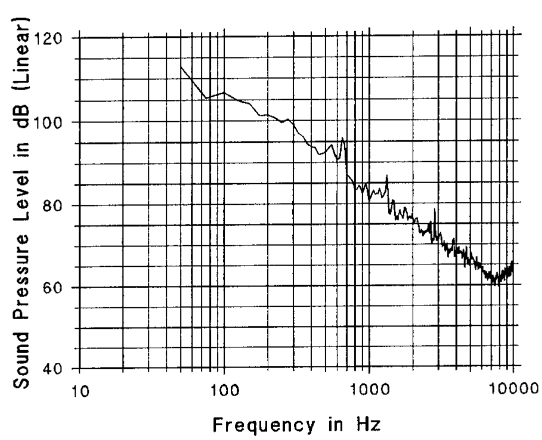


Figure 6. S-61. Noise level in the pilot's seat.

In the *Westland Lynx helicopter*, the noise impact is lower but dominated by the very irritating peaks at approximately 1 kHz, caused by the gear box (please, see fig. 7 next page).

In conclusion our recordings demonstrated the need for efficient low frequency protection in propeller driven aircraft and the need for more efficient protection in helicopters.

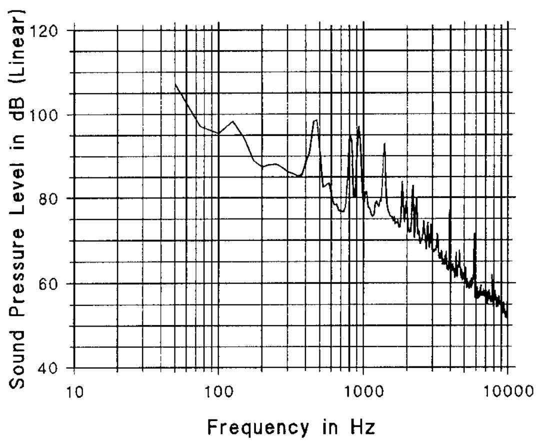


Figure 7. Lynx noise spectrum in pilot's seat, the helicopter cruising at 100 knts.

Noise protection devices:
The collection of helmets and head-sets tested were the following:

1. Astrocom head-set used in the C-130
2. BOSE head-set making use of active noise reduction (ANR) priciple.
3. HGU-55 helmet,
4. SPH-3 helmet,
5. HGU-84.

All helmets are custom fit to the test person in question.

The noise attenuating properties of the Astrocom head-set appears from fig. 7. 20 dB at 1 kHz, but virtually nothing in the low frequencies.

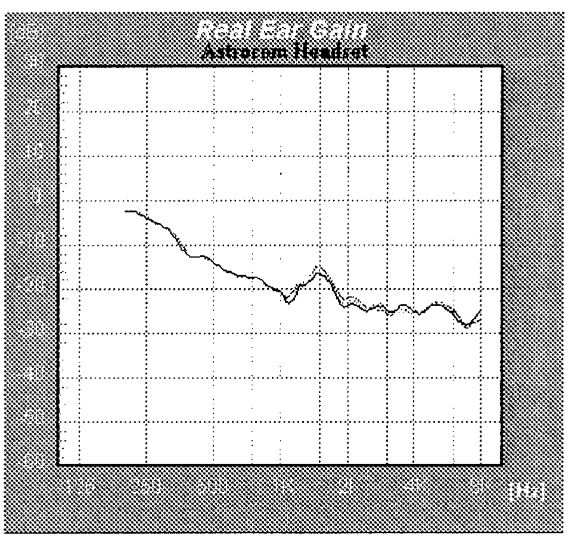


Figure 8. Noise protection properties of the Astrocom head-set.

If a self expanding silicone ear plug is inserted in the ear canal before putting the head-set on, the attenuation improves significantly as expected. This is obvious in the low frequencies but also indisputable at higher frequencies (fig. 9, below).

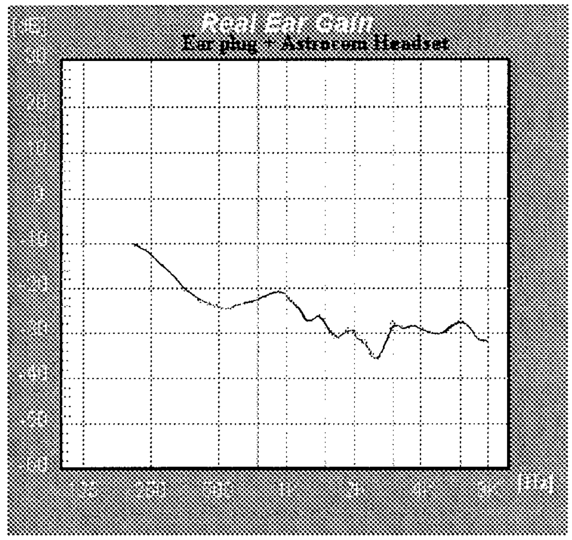


Figure 9. Noise protection properties of the Astrocom head-set combined with the use of a simple ear plug.

The SPH-3 helmet provides a noise protection very much like the protection produced by the Astrocom head-set plus the ear plug. Low frequencies are significantly attenuated. It is very efficient when combined with the use of an ear plug (fig. 10).

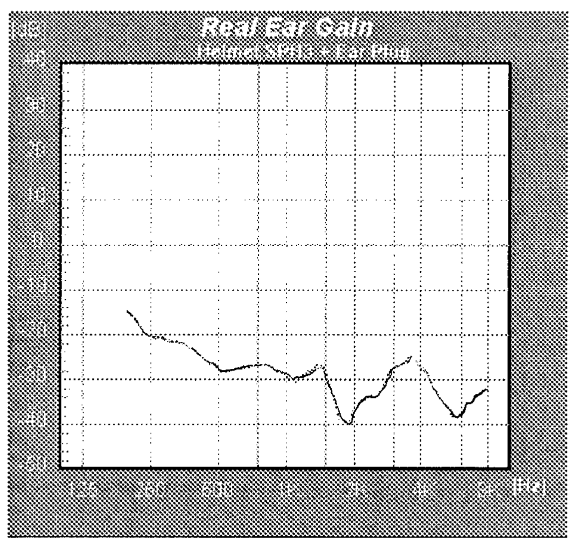


Figure 10. SPH-3 helmet combined with an ear plug.

Neither the HGU-55 (fig. 11, next page), nor the HGU-84 (fig. 12, next page) are very efficient protectors against low frequency noise.

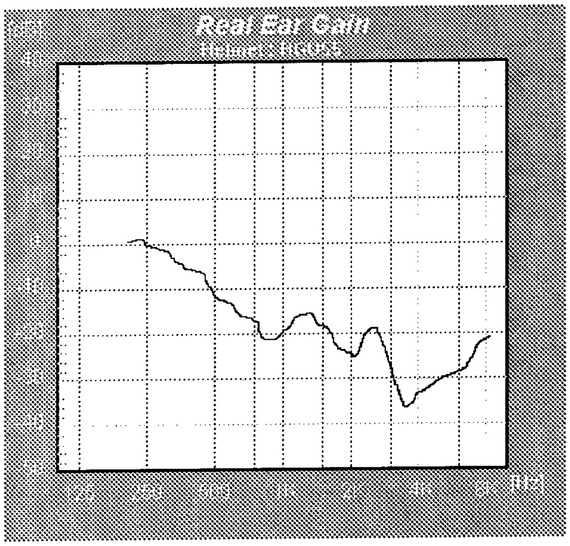


Figure 11. The HGU-55 helmet.

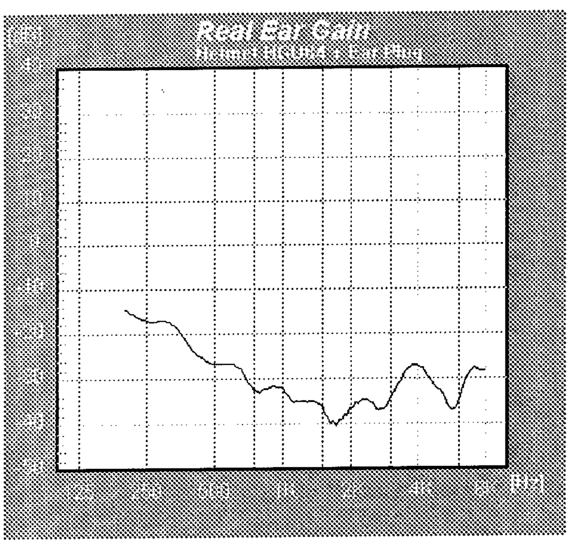


Figure 12. Noise reduction properties of the HGU-84 helmet when combined with an ear plug.

The ANR BOSE head-set provides a new perspective in noise protection by actively counteracting the noise. The efficacy of the active mechanism was studied by simply switching on and off the electronic box.

It's obvious from fig. 13 that the active mechanism works - but most efficiently in the low frequencies with no effect what-so-ever in the mid- and high frequencies.

When combined with the use of an ear plug, the frequency distribution seemed more attractive, but still most efficient in the low frequencies.

In order to further evaluate the efficiency of the BOSE

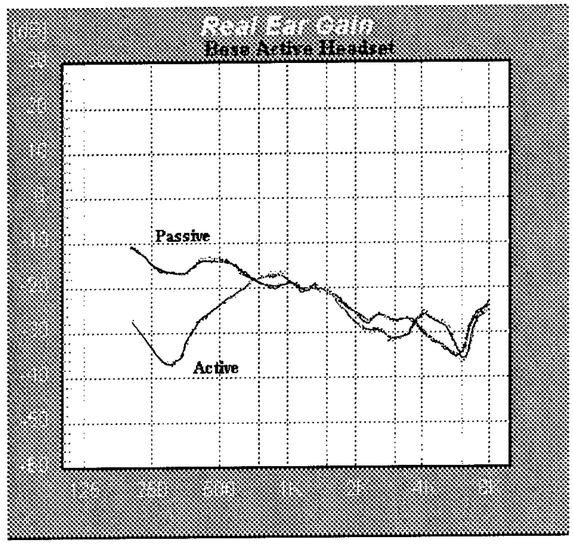


Figure 13. The efficacy of the active mechanism of the Bose ANR head-set.

head-set, we tested 10 normally hearing subjects in a C-130 noise environment. Our standard word list used for determination of the threshold of intelligibility during routine speech audiometry was presented to the subjects mounted with the head-set. Two different situations were evaluated - the mechanism on and the mechanism off. The results appears from fig. 14.

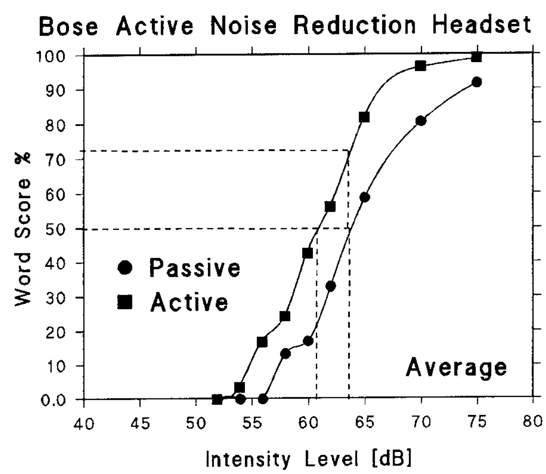


Figure 14. The efficacy of the active mechanism of the Bose ANR head-set, evaluated as improvement of speech intelligibility.

The difference in performance seems very small, 3 dB. But intersecting the 50% level of the off-condition with the on-condition curve shows that a gain of approximately 22% in intelligibility can be expected. What may be more important is that all subject found the on-condition far less annoying than the off-condition.

CONCLUSION:

Our study confirms the well-known fact, that low-frequency noise predominates in propeller-driven aircraft, including helicopters and consequently there is a need for noise protection devices efficient in the low frequency range when flying that type of aircraft. A significant extra low-frequency attenuation can be obtained by the use of a self expanding ear plug when wearing a helmet or a head-set. Indeed, this both reduces any possible signals and the noise.

The active noise reduction (ANR) head-set (BOSE) tested is only efficient below 750 Hz and seems more attractive in use than ear plugs. The effects on speech intelligibility of the active mechanism of the ANR seems not to be very dramatic, but all subjects tested found that using the ANR resulted in a much more attractive acoustic environment.

References:

1. Burns, W. Noise and Man. Chapter 13: Aircraft Noise. John Murray, 50 Albermarle Street, London, 1968, p. 206.

2. International Standard Organization. Acoustics - Measurements of Noise Inside Aircraft. ISO 5129:1989.

Coordinated Speech Technology Research of nine NATO countries in Research Study Group RSG.10

H.J.M. Steeneken* and L. Gagnon**
* TNO Human Factors Research Institute, P.O. Box 23, 3769 ZG Soesterberg, the Netherlands
** Department National Defence, P.O. Box 9703, Ottawa, Canada K1G 3Z4

1. SUMMARY

The relevance of the application of speech and language technology in the military is considered. The Research Study Group initiates and coordinates research on these applications, focused on the specific adverse military conditions. Specifically multi-lingual applications with non-native speakers, adverse environmental conditions (high noise, g-forces, vibration), and stress conditions (workload, battlefield) do reduce the performance of advanced applications. For example secure voice at low bit-rates, speech recognition in command and control, speaker/language recognition for intelligence and translation in joint force conditions are major applications to be studied. Nine NATO countries participate in this effort.

2. INTRODUCTION

Speech communication between humans and with computing systems or via communication systems is recognized as an important facility in command and control, aircraft and vehicle operations, military communication, information retrieval, intelligence, training, and voice surveillance. Since its establishment in 1977 the NATO research study group on speech processing (AC243(Panel 3)RSG.10) conducts experiments and surveys focused on these subjects in military applications.

Presently nine countries are represented in RSG.10: Belgium, Canada, France, Germany, The Netherlands, Portugal, Spain, United Kingdom, and United States of America. The group initiates and coordinates studies, experiments, and surveys on speech and language technology focused on the specific military requirements. In many military applications in the multinational NATO community, adverse conditions are to be dealt with. High background noise levels, vibration, g-forces, (battlefield) stress, limited quality of transmission systems, and non-native talkers and listeners make use of speech as a means of communications between humans and with machines difficult. Present state-of-the art technology is in general developed for civil applications and not developed for adverse conditions. Therefore, the majority of the activities of RSG.10 are related to improvement of the performance of systems and technologies for the military. Tools for development and evaluation such as specific data bases were collected and disseminated. For example: a noise data-base with 24 representative military noises, and a multi-lingual isolated and connected spoken digit data base including non-native speakers were produced.

In this paper a few recent projects are discussed.

3. RECENT AND PRESENT COORDINATED PROJECTS

Advances in the relevant speech processing techniques have been substantial and their application has been increasing steadily over the last three decades. This has several causes:

- The heavy workload of personnel working in command and control environments, avionics and operator positions can be substantially reduced by using voice input and/or output systems.
- In many situations operators have busy eyes and hands, and must use voice for control functions.
- Modern training environments are supervised by a system rather than by instruction personnel (e.g. air traffic controllers). For interaction with the student advanced voice input/output systems are required.
- For information retrieval large vocabulary speech recognition and speech understanding systems are being developed for applications as mission preparations and battlefield services.
- Speech processing techniques allow secure voice communication at very low bit rates.
- Speech processing techniques are being developed to permit the identification of talkers, of language spoken, and of keywords in broadcast communications.
- Worldwide operation of military units requires for intelligence purposes advanced speech translation and understanding systems.

The use of these systems in a military, multi-lingual, environment requires specification and assessment of methodologies specially adapted for these purposes. National research programs often concentrate on processes specific to the national language. Cooperative research projects across NATO nations have served to extend the techniques to multi-lingual environments. Continuing exchange of information between NATO nations, cooperative development of assessment methods and identification of future military applications are therefore highly desirable.

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

Recent cooperative studies were focused on:
- automatic speech recognition in combination with high environmental noise,
- the effect of mental stress on the production of speech and the performance of systems,
- the potentials of speech and language technology for military applications.

Noise studies

A study on the effect of noise on the performance of automatic speech recognition systems was carried out with two objectives: (1) to address some of the problems of assessment in a standardized and reproducible manner and (2) to provide a calibrated data base for further assessment purposes (phase 1).

With this data base various state-of-the-art systems were evaluated in order to obtain information on the applicability of these systems in the military (phase 2).

The first activity was to compare the results under identical conditions reproduced at different sites (standardization through the use of common data, careful calibration and the use of a reference recognizer) and to stimulate dialogue on future assessment work in this area (e.g. on suitable future databases and experiments). To this end the RSG.10 laboratories have used the existing RSG.10 noise and digits databases and (in liaison with the ESPRIT SAM Project laboratories) have developed the NOISEX-92 experiment and CD-ROMs (Steeneken and Varga, 1993).

NOISEX-92 is intended to provide a common database with which to work in the short term as well as to highlight future requirements. It does not address the full range of factors affecting speech recognition in noisy environments; the data are for instance artificial in nature because the speech and noise have been separately recorded and added together arithmetically. NOISEX-92 does, however, provide a set of control data that can be easily used.

With this data base phase 2 of the project was conducted (Gagnon and Cupples 1995) in which five laboratories participated. Evaluation of the effects of different types of authentic noises on the recognition performance was performed for several modern automatic speech recognition (ASR) systems. Also the ability of various types of noise compensation techniques to improve recognition performance was studied. The results obtained from five laboratories show that ASR systems alone are not noise robust and that noise compensation techniques can extend the range of operation and potential applications.

Effect of stress on speech production

A study on the effect of stress on speech technology systems started in 1995. Military operations are often conducted under conditions of stress, induced by high workload, sleep deprivation, and battle stress. These stresses are believed to affect voice quality, and are likely to be detrimental to the performance of communication equipment (e.g., low bit-rate secure voice systems) and weaponry with vocal interfaces (e.g., advanced cockpits, command and control systems). The actual effects of stress on voice are not well understood. RSG.10 has conducted a survey of the literature on stressed voice, and has concluded that it is necessary to conduct a study of stress effects of the kind to which military operations are subject. Only with the results of such studies it will be possible to assure the performance of vocal systems under operational conditions.

An additional result in these studies is likely to be the possibility of assessing the stress on personnel during operations by analysis of the voice. The study, however, will not be primarily directed to this end.

The work is subdivided into six tasks:
(1) Collect data with various types of stress, such as workload. Physiological measures correlated with other objective measures will be collected in parallel,
(2) Produce an annotated database that might be used beyond the confines of RSG.10 (continuous through the life of the project),
(3) Characterize speech parameters related to stress,
(4) Assess effects on performance of recognizers and communication equipment,
(5) Workshop provision of database for analyses,
(6) Military applications of derived results.

A workshop open to the international research community has been held already (Moore and Trancoso, 1995). Also a database has been collected consisting of speech recorded under various stress conditions as sleep deprivation (contribution of DCIEM Canada), air traffic control, and aircraft in crash condition.

Presently the analysis of these speech data is in progress. The study will also focus on the effect of this type of speech on the performance of narrow-band voice coding systems, speech recognition systems, and speaker recognition (intelligence tasks).

Study on the potentials of speech and language technology in military applications

The key military application areas for speech and language technology, as indicated in Table I, are: command and control; communications; computers and information access; intelligence, training; and joint (coalition) forces. In general, all the speech and language technology areas have some application to all the application areas, but Table I highlights particularly important connections between applications and technologies.

For each category a description of the requirements and possible goals is given. The available technologies are subdivided in:

- Speech Processing
- Language Processing
- Interaction
- Assessment and Evaluation.

For these technologies the state-of-the-art with respect to performance and availability is discussed. For speech processing a sub-division for speech coding, speech synthesis and recognition is made.

Also an overview is given of possible assessment procedures and design criteria. The study also included some case studies and applications.

In brief the report highlights the need of speech control for operational systems and advanced communications in a changing military environment. Reduction of personnel, increasing complexity of systems, multi-national operations require optimal human performance in which speech can be a natural means of interfacing.

Table I. Overview of areas of military applications of speech and language processing in relation to available technologies. The numbers refer to the corresponding paragraphs in Steeneken et al. 1996.

Available Technologies / Speech and Language in Military Applications		3.1 Speech Processing	3.1.1 Speech Coding	3.1.2 Speech Enhancement	3.1.3 Speech Synthesis	3.1.4 Speech Recognition	3.1.5 Speaker Recognition	3.1.6 Language Identification	3.2 Language Processing	3.2.1 Topic spotting	3.2.2 Translation	3.2.3 Understanding	3.3 Interaction	3.3.1 Interactive dialogue	3.3.2 Multi-model communication	3.3.3 3-D Sound Display
Command and Control	2.1		•	•	•	•						•		•	•	•
Communications	2.2		•	•							•					•
Computers and Information Access	2.3				•	•					•	•		•		
Intelligence	2.4		•	•		•	•	•		•	•	•		•		•
Training	2.5				•	•								•	•	•
Joint Forces	2.6		•	•							•	•				
Case studies	5.0															
Cockpit Fast jet	5.1		•	•	•	•									•	•
Helicopter	5.2		•	•	•	•									•	•
Sonar	5.3					•										•
Noise reduction	5.4		•	•												
Training of air traffic controllers	5.5				•	•								•		•
Spoken Language Systems demonstration	5.6				•	•		•			•	•		•	•	
Voice Technology in Space	5.7				•	•				•				•	•	•
Speech Coders 600-1200 Bps	5.8		•	•												

4. CONCLUSION

Exchange of information, cooperation in studies focused on the specific military requirements for speech technology products is the main goal of the NATO research study group on speech processing (AC243(Panel 3)RSG.10). The participation of nine countries indicates the interest of the various nations on this topic.

It is high lighted that speech and language are major means of communication. Also command and control and training have a need for advanced human friendly interfacing with systems or training equipment.

The multi-lingual NATO community requires additional attention to specific factors with respect to speech technology. Non-native speakers, adverse environmental conditions, and stress conditions require a more robust technology than used in identical civil applications.

The RSG.10 is focused on these topics and has conducted and initiated many international projects. This

may be reflected by the condensed literature overview given below.

Presently a number of organizations are supporting international projects on speech and language (European Union DGXIII, ARPA) or are organizing international conferences and workshops (ESCA, IEEE). However, the specific mission of RSG.10, focused on the specific military requirements, is not covered by these bodies. RSG.10 encourages however cooperation such as performed in the past with the joined organization of five workshops with ESCA.

5. REFERENCES

Vonussa, R.S., Nelson, J.T., Smith, S.E., and Parker, J.G. (**1982**). NATO AC243(Panel 3) RSG.10 Language data base. Proc. US National Bureau of Standards workshop on standardization for speech I/O Technology Gaithersburg, pp. 223-228.

Pols, L.C.W. (**1982**). "How Humans perform on a Connected Digit Database", Proc. IEEE ICASSP Paris (867-870).

Bridle, J.S. (**1983**). "Connected word recognition for use in Military systems", NATO DRG document AC243(Panel 3)D854 and D128.

Hunt, M.J. (**1983**). "The Speech signal", Proc. NATO AGARD Lecture series No. 129, Speech Processing. pp 2.1-2.12

Hunt, M.J. (**1983**). "The Speaker differences in Speech and Speaker recognition", Proc. NATO AGARD Lecture series No. 129, Speech Processing. pp 5.1-5.8.

AGARD lecture series No. 129 (**1984**). "Speech Processing", Ten selected papers by members of RSG.10. AGARD Neuilly sur Seine, France.

Steeneken, H.J.M. (**1985**). "Recommendations for analogue tape-recording of speech signals for data bases", NATO DRG document AC/243(Panel 3)/D8.

Cote, D.O., Kaufmann, R., and Taylor, M.M. (**1985**). "Voice versus keyboard for database Query", J. Amer. Voice Input-Output Soc., 2, pp.1-22.

Hunt, M.J. (**1986**). "Speech is more than just an Audible Version of Text" AGARD Conference Proceedings No 414: Information Management and Decision Making in Advanced Airborne Weapon Systems Toronto, pp.22.1–22.12.

Taylor, M.M. (**1986**). "Issues in the evaluation of Speech recognition systems", J. Amer. Voice Input-Output Soc., 3, (39-68).

Moore, R.K. (**1986**). The NATO research Study Group on Speech Processing: RSG.10. Proc. Speech Techn., New York, pp.201-203.

Steeneken, H.J.M., and Houtgast, T. (**1986**). "Comparison of some methods for measuring speech levels", Report IZF 1986-20, TNO Institute for Perception, Soesterberg, The Netherlands.

Moore, R.K. (**1988**). "Connected Speech Recog-nition in a multi-lingual environment", NATO DRG-document AC243(Panel 3)D11.

Steeneken, H.J.M., and Geurtsen, F.W.M., (**1988**). "Description of the RSG.10 Noise data-base", Report IZF 1988-3, TNO Institute for Perception, Soesterberg, The Netherlands.

Moore, R.K. (**1987**). "Speech Technology in a multilingual NATO environment", Proc. Mil. Speech Techn., Washington.

Taylor, M.M., and Hunt, M.J. (**1988**). "Flexibility versus Formality", in *Structure of Multimodal Dialogue*, M.M. Taylor, F. Néel, and D. Bouwhuis eds.., Amsterdam: North Holland, pp.435-454.

Taylor, M.M., Néel, F., and Bouwhuis, D.G. (**1989**). "Structure of Multi modal Dialogue", Amsterdam: Elseviers Science Publishers (North Holland).

AGARD Lecture series No. 170 (**1990**). "Speech analysis and Synthesis and Man-Machine Speech Communications for Air Operations", Eight selected tutorials by members of RSG.10. AGARD Neuilly sur Seine, France.

Hunt, M.J. (**1990**). "The Speech Signal" Proc. NATO AGARD Lecture series No. 170. pp 2.1 - 2.12. Also in *Digital Speech Processing: Speech Coding, Synthesys and Recognition,* Kluwer Academic Publishers, Boston, (**1991**), ed. A. Nejat Ince, Chapter 2, pp 43-71.

Hunt, M.J. (**1990**). "Figures of Merit for Assessing Connected-Word Recognisers", *Speech Communication*, Vol 9, pp. 329-336.

Weinstein, C.J. (**1991**). "Opportunities for advanced Speech Processing in Military Computer-based Systems", NATO-DRG report, also Proc. IEEE (**79**) 1626-1641.

Varga, A., Steeneken, H.J.M., Tomlinson, M., and Jones, D. (**1992**). "The NOISEX-92 study on the effect of additive noise on automatic speech recognition", Description of RSG.10 and Esprit SAM experiment and data base, DRA Speech Research Unit, Malvern, England.

Varga, A., and Steeneken, H.J.M. (**1993**). "Assessment for automatic speech recognition: II. NOISEX-92. A database and an experiment to study the effect of additive noise on speech recognition systems", Speech Communication, (**12**), pp.247-252.

Mangold, H., Hunt, M., and Néel, F. (**1993**). "Applications of Speech Technology", Proceedings of NATO/ESCA workshop Lautrach, NATO DRG document AC243 (Panel 3)TP/4.

Steeneken, H.J.M., and Varga, A. (**1993**). "Assessment for automatic speech recognition: I. Comparison of assessment Methods", Speech Communication, (**12**), pp.241-246.

Steeneken, H.J.M., and Varga, A.P. (**1993**). "Automatic Speech Recognition in Additive Noise (I): (pilot study and the calibration and standardization and the use of common data bases)", NATO DRG document AC243 (Panel 3)TR/11.

Taylor, M.M., Néel, F., and Bouwhuis, D.G. (**1994**). "Structure of Multi modal Dialogue II", in preparation.

Gagnon, L., and Cupples, E.J. (**1995**). "Automatic Speech Recognition in Additive Noise", NATO DRG document AC243(Panel 3)TR/17.

Moore, R.K., and Trancoso, I. eds. (**1995**). Proceedings Workshop on "Speech under Stress", NATO/ESCA workshop, Lisbon Portugal, also as NATO DRG document AC243(Panel 3)TP/11.

Steeneken, H.J.M. ed. (**1996**). "Potentials of Speech and Language Technology Systems for Military Use: an Application and Technology Oriented Study", NATO DRG document AC243 (Panel 3)TR/21.

Speech language hearing test results of active duty pilots failing the pure tone audiometry limits of ICAO guidelines—Method , Problems and Limits to verify the waiver status

W. Hanschke
German Air Force Institute of Aviation Medicine, Division I - Medical Examination and Consulting Service - Otorhinolaryngology Department,
P.O. Box 1264 -KFL, D-82242 Fürstenfeldbruck, Germany

Summary:

Adequate hearing is essential for communication in flight and rapid and accurate assessment of warning tones in the cockpit. Waiver is permitted, when hearing is adequate to permit essential communication in flight. The Freiburger speech language hearing test method gives the opportunity to verify the intelligibility in a standard proven manor with the possibility to add aviation related necessities. A higher safety standard could be refined by replacement of the forner subjective aeromedical hearing methods.

Introduction:

The German military guidelines for pure tone audiometry have been discussed. ICAO guidelines for audiometry threshold were less strict than German military guidelines and those used in German occupational medicine for assessment and recommendation for workers in noise environment. The test method should allow to give fair recommendation and to be safe in legal aspects. Therefore the differences between the clinical and aerospace medicine methods were not understandable. While recommendations and test batteries in clinical and occupational medicine could be used to quantify hearing losses and had been accepted as standards, when used by expert witnesses for disabling determinations in German courts—the aerospace medical tests were mainly based on subjective impressions and couldn't be quantified to be accepted as standards in German courts.

The aeromedical concerns are :

„Adequate hearing is essential for communication in flight and also for rapid and accurate assessment of warning tones in the cockpit."

Problem description:

The hearing standards are varying between the different Nato forces as well as the civilian national guidelines. Meanwhile the German military and civilian standards for pure tone audiometry limits for unaided hearing do not differ for active duty professional pilots. The differences to the USAF military standards as well as other NATO-countries standards are minor, but we have to face in multinational units the differences and to accept them.

Pure tone audiometry standards: maximum unaided hearing loss in dB(A)

	250Hz	500Hz	1KHz	2KHz	3KHz	4KHz - 6KHz
WFV I	20dB	20dB	20dB	20dB	**max 25dB per frequency**	
WFV II&III	30dB	35dB	35dB	35dB	50dB	**no recommendations**
civilian I&II	-----	35dB	35dB	35dB	50dB	**no recommendations**
civilian III	**no audiometry, speech intelligibility distance 2m normal spoken language both ear hearing**					

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

USAF -- AFPAM 48 –132 attachment 10 - 12. Aug. 93

H1	-----	25dB	25dB	25dB	**sum less than 270 both ears at 3, 4, 6 KHz**
H2 -better ear	----	30dB	30dB	30dB	**no recommendations**
H 2-worse ear	----	30dB	50dB	50dB	**no recommendations**

Perhaps research in the future will equalize the standards in pure tone audiometry. But what do we have to do with those, who failed the pure tone audiometry standards?_Military as well as civilian guidelines do not describe a speech language hearing test with limitations as we know it for the pure tone audiometry.

The military advice about speech language hearing test: If the pilot fails the limits, a sufficient intelligibility has to be verified by speech language hearing tests in an anechoic chamber as well as under usual cockpit noise levels. **The civilian guidelines:** If there is flight experience ... intelligibility testing under cockpit noise level conditions should be done**.** -- Nothing is said about the limits! How much is necessary? How do we understand sufficient speech intelligibility by means of speech language hearing tests?

- What are the limitations - 80% - 90% - 100%?
- Which test should be used to verify the results?
- Should speech intelligibility be tested in the native language (mother tongue) and do we consider these results could be transferred to the most common aerospace language - (English in all the different ways of pronouncing round the world?
- How loud should the radio power be set?
- Could we allow flying in noise environment and radio power set above the limitation as far as the pilot needs it for intelligibility?

In the past adequate hearing after failing the pure tone audiometry standards was verified by a sentence comprehension test in an anechoic chamber as well as inflight hearing tests under cockpit noise levels. **We can follow a conversation in our mother tongue, even when we couldn't under-stand nearly 50% of the spoken words in a quiet surrounding. This intelligibility will drop rapidly under noise environment!!!**

- But what is with those pilots, who have to speak a foreign language in a noise environment?
- What are their limitations flying in Europe, where there are nearly 50!!! (fifty) official languages and only one of these is English - different from Scottish, Welsh or Irish accent?
- What are their limitations flying round the world at areas with pure English pronunciation**?**

Knowing how difficult it is to follow different mother tongue speakers in their pronounciation, we have to consider this question and to decide about the sufficiency of our test methods.

Often the hearing and the intelligibility may be even enough for the daily work, but when being appointed to multinational organisations most of our borderline handicapped officers noticed their crippled hearing and asked for hearing aids. Sometimes they came with all kinds of symptoms.

Often an audiometry test and the following anamnesis showed the real reasons. Mostly they are afraid to wear the hearing aids, because of a crippled stigma, which may cause some feelings of resentments. This we have had to face after the reunification of Germany, when we had to refine aeromedical hearing standards, because of more severe hearing losses among Eastern German Pilots as well as some recent cases of hearing loss among older West German pilots (mainly rotary pilots). The legal side of disqualification in the difficult political situation had to be considered as well.

Methods and Limitation:

The **German Freiburger Word Discrimination Test** seemed to be a fair test method, because this test has been used in German occupational medicine to quantify hearing loss and had been ac-cepted as a standard, when used by expert witnesses for disabling determination in German courts. **The subjective impressions could be replaced by objective measurements and proven tests. The German Freiburger Word Discrimination Test is based on two components:** The understanding of polysyllabic numbers - with with evaluating the 50% hearing level.The understanding of monosyllabic words at 60-80-100dB -20 words were tested at each decibel (A) level. These components are administered in a standardized manor and the results have been proven repro-ducible, when used by civilian medical expert witnesses. Less than 80% discrimination at 80dB(A) disqualifies for further flying duty! ***For military necessities language communication must be possible without missunderstanding in high noise environments.*** German military demands are: ***The high level of flight safety means to deal with the worst case scenario - adequate hearing must be possible on the worse ear, because it could be the remaining hearing ear, if there are radio, headset or speaker problems!*** The pilot should understand and fly safely under the worst possible conditions and we as flight surgeons do not want to add a known handicap to an unknown situation and give responsibility back to the pilot, who knows about his handicap and thinks about the influence of his performing. **He would never fly with technical equipment, which has the same percentage of malfunction as we could measure his percentage of hearing loss.** Therefore language discrimination was tested also by masking with 80 dB and 100dB white noise one ear and testing the intelligibility of language communication at **80dB(A).** A waiver is recommended, if less than 5% decrements in intelligibility is verified in relation to the standard testing in the hearing booth.

Speech language audiometry hearing booth

- 80dB - speech language hearing - **result**: 95% correct
- 100dB - speech language hearing - **result**: same or more, - if less the test is failed because of medical reasons

Cockpit noise simulated Speech language audiometry

- Tested ear: 80dB - speech language level
- Opposite ear: 1. 80dB and 2. 100dB white noise - in the future real cockpit noise

Result must be the same compared to hearing booth conditions !

The early results of this test as well as the pilot impression's meant that they feel more safe even if they have their

hearing problems during party noise levels. Having dealt now with the hearing conditions I would like to introduce another aspect of the Freiburger speech language hearing test, which perhaps is an unknown possibility and chance.The Freiburger Speech language test could be used to verify the pronunciation and ability to understand the pilots spoken words as well, if you document the spoken words with a voice recorder and give it to different secretaries to write.
Additionally the pronunciation under real noise frequencies must be documented by protocol and the failure must be less than 5 %. To test pronunciation is necessary in such cases as major cancer surgery, vocal cord dysfunction, dental problems in connection with speech or facial nerve dysfunction. For language hearing discrimination the Freiburger language hearing test is standardized, for pronunciation the evaluation is in the early beginning and a 100% is recommended.
Up to now I have only experience in testing two cases and the responsible commanding officer , the flight safety officer as well as the responsible flight surgeon felt satisfaction during his procedure and the following flying status.

- **References:**

von Werra, P.; Tschopp, K.; Schneider, R.
Einfluß der linguistischen Kompetenz auf die sprachaudiometrischen Ergebnisse mit dem Baseler Satztest
HNO (1995) 43;98-103 - Springer Verlag 95

2. Apter-Yehezkely, G.; Muchnic; H.; Feinmann, R.; Schamiss, A.; Davidson, B.
Evalution of speech discrimination in sensory-neural High-Tone-Loss Air Crew

3. Becker,W.; Naumann,H.-W.; Pfaltz,C.-R.
Hals-Nasen-Ohrenheilkunde
Thieme, 3. Auflage 1986

4. Fleischer, K.; Kießling,J.;
Topographische Audiometrie
HNO Praxis Heute 1, Springer 1980 (S.1-20)

5. Niemeyer, W.
Aggravationsprüfungen und objektive Audiometrie
HNO Praxis heute 4, Springer 1984 (S.1-38)

6. Lehnhardt, E.
Praktische Audiometrie, 5. Aufl., 1978
Thieme Verlag

7. Brusis, T.
Die Lärmschwerhörigkeit und Ihre Begutachtung
Demeter Verlag 1978

8. Plath, P.
Lärmschäden des Gehörs und Ihre Begutachtung
Schriftenreihe HNO - Kind Hörgeräte1991

10. Feldmann, H.
Das Gutachten des HNO-Arztes, 3. Auflage
Thieme 1994

11. Hahlbrck, K.-H.
Sprachaudiometrie, 2. Auflage
Thieme 1970

12. Fröhlich, G.
Tongehör des altenden Flugzeugführers
Arbeitsmedizin/Sozialmedizin 7, 1972, S. 261-266

13. Fröhlich, G.
Die Auswirkung von Hörverlusten im Tonaudiogramm auf das satzverständnis von Flugzeugführern in Ruhe und im Lärm
Wehrmedizinische Monatszeitschrift März 1976

14. Fröhlich, G.
Fröhlich,G.
Lärmbedingte Hörstörungen und ihre Auswirkung auf das Sprachgehör
Kassenarzt 17, 1977, S. 4432-4434

15. Fröhlich, G.
Der Einfluß von Gegörschutzgeräten und Hörverlusten auf die Sprachverständlichkeit im Fluglärm
Zentralblatt Arbeitsmedizin 1978 / 3, Bd. 28

Improved Speech Intelligibility in Aircraft Noise due to Altitude

Anthony S. Wagstaff, M.D.

Royal Norwegian Air Force Institute of Aviation Medicine
PO Box 14, Blindern, 0313 Oslo, Norway

Summary

Few studies have addressed effects of altitude and noise combined, although these two effects are inherent parts of all aviation. The few published studies that have addressed altitude effects on hearing function have mainly focused on using gas mixtures, and have demonstrated inconclusive results. The present study was designed to assess the effect of altitude on speech intelligibility in aircraft noise. The primary hypothesis was a predicted detrimental, hypoxic effect on speech intelligibility in noise. Eight male subjects with normal hearing were fitted with an aviation headset specially adapted for use with the audiometer. Pure-tone audiometry, as well as speech audiometry in noise, was performed at 0, 10.000, 13.000, and 16.000 ft. simulated altitudes in a hypobaric chamber. The 4 test altitudes were performed double blind with respect to audiometry operator and test subject. Arterial blood gases where measured using an intra-arterial catheter and tympanometric measurements verified full middle ear equilibration. Noise levels where monitored and logged throughout all experiments.

A substantial increase in speech intelligibility in noise due to altitude was found in this study. The physical effect of barometric pressure on noise causing an increased signal-to-noise ratio was found to greatly outweigh any hypoxic detrimental effect.

Introduction

The effects of altitude, such as hypoxia and decreased barometric pressure, have always been very central in the field of Aviation Medicine. Noise is also an important environmental factor which is always present in aviation. However, much is still not clear regarding the effects of these environmental stressors combined. Audition is, arguably, the most important sense in flight apart from vision. In addition, the relative importance of the auditory sense is definitely increasing. This is due not only to the increasing task complexity and communication environment in an ever-increasing traffic picture on the Civilian side. In Military aviation, as well as in Civil aviation, there is a growing awareness of audition as a sense that has not been fully utilised. The potential of improving and increasing information via the auditory sense is exemplified by recent research into three-dimensional auditory displays (1).

The human auditory system is obviously inherently dependent on a stimulus to function, likewise is any incoming stimulus dependent on propagation through whichever medium surrounds us. The vast majority of research into human audition has been performed in normal atmospheric conditions. When man is put into an aviation environment of reduced atmospheric pressure, important changes take place which might have an impact on auditory perception. Two main factors are usually mentioned, the pressure change itself, the other being the resulting hypoxia due to a reduced partial pressure of oxygen. These factors have to some degree been studied as separate factors. However, in Aviation, they almost always occur together. Some of the research that has been performed to evaluate hypoxia and pressure effects respectively, is mentioned in the following:

One of the procedures for studying the effects of hypoxia on hearing has been using gas mixtures with an increased nitrogen content. Smith and Seitz published an article in 1946 using this technique (2), where a

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

decrease in speech intelligibility was reported at an oxygen tension corresponding to 16,900 feet altitude. Smith published another study in the same year (3), using the same technique, where a prolonged exposure (>6 hrs.) to a simulated altitude of 10,000 feet had no reported effect on speech intelligibility. Similarly, pure tone thresholds were investigated by Klein et al. in 1961 (4). This group of workers found that subjects inhaling a Nitrogen-Oxygen mixture roughly equivalent of an altitude of 20.000 feet, showed a decrement in hearing sensitivity at lower/middle frequencies, and a slightly improved sensitivity at 4096 Hz. Klein (5) also published a paper showing similar effects on bone-conducted thresholds in a noisy environment. Several possible explanations for these results were offered, but the validity of the above experiments to the aviation environment is limited. The use of a different gas than air may affect sound propagation, and reduced atmospheric pressure is an integral part of an aviation environment producing hypoxia.

There are not many published articles regarding audition and hypoxia as produced in an altitude chamber. This may be due to the inherent methodological problems, since all standardisation and calibration of auditory measuring techniques relates to normobaric air. Curry and Boys (6) published a study in 1956, where they reported no change in pure-tone thresholds at a simulated altitude of 15,000 feet. This paper gives a good overview regarding some of the methodological problems that were perceived in relation to their experiments.
Burkitt and Perrin published a study in 1976 (7) where both thresholds for pure tones and speech were measured at 15,000 and 20,000 feet simulated altitudes in an altitude chamber. They found no significant change in pure-tone threshold at these altitudes, but a significant detrimental effect on speech intelligibility was reported. This effect was attributed to Central Nervous System hypoxia.

Recent work on Auditory Reaction Time and P300 Latency (8,9), show a clear effect of gas-mixture induced hypoxia at 65% arterial oxyhaemoglobin saturation on both these parameters. This was reportedly equivalent to an altitude of about 4200 m (13800 feet). The point was made that audition may be more sensitive to hypoxia than is currently believed.

Although many of the above mentioned papers indicate some hypoxic effect on auditory parameters from around 14,000 feet simulated altitude or thereabouts, several points prevent us from extrapolating the above results into the aviation environment. Nearly all experiments have been performed in almost silent surroundings, and mostly inducing hypoxia with Nitrogen/Oxygen gas mixtures. The findings of different workers vary substantially, particularly in relation to pure-tone thresholds. Thus, little is still known regarding how altitude affects the communication performance at the receiving end in the real aviation environment.

The present study aims at investigating the effect of altitude on speech understanding in realistic aircraft noise, thus, providing results which as far as possible can be applied to the operational aviation environment. The study has therefore been designed to answer this question in an applied fashion, rather than to investigate hypoxia and pressure effects separately.

Materials and Methods

Subjects

Eight male subjects 25-33 years of age completed this experiment. All had hearing thresholds less than 20 dB for the frequencies 250-4000 Hz. For frequencies of 6000 and 8000 Hz, up to 30 dB thresholds were allowed for single frequencies. All subjects were otologically normal, with no history of serious ear disease or any other predisposing conditions. Otological examination was performed, and was normal for all subjects. Likewise, tympanometry was performed since failure to equilibrate pressure has been shown to affect hearing (10).One subject was not allowed to participate due to an abnormal tympanogram, and one subject had to discontinue the experiment due to severe discomfort and faintness at simulated altitude. The eight subjects who completed the whole experiment tolerated it well.

The experiments were approved by the Regional committee for Medical Research Ethics, and written informed consent was likewise obtained from all subjects.

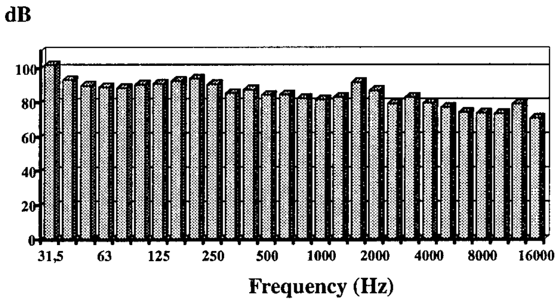
Experimental environment

The present experimental series was conducted in the Altitude chamber facility at the the Royal Norwegian Air Force Institute of Aviation Medicine. This altitude chamber has an interior similar to an aircraft cabin and is well insulated, providing good sound proofing. Temperature and humidity was kept within small variations for all altitudes, the temperature being kept at $25^{0}C$ +/-$1^{0}C$, with two short excursions up to $28^{0}C$ in one experiment.

Ambient noise was produced by 2 speaker systems placed at one end of the chamber. Apart from the power amplifier, all the other playback equipment was situated outside the altitude chamber. The subject was placed facing the loudspeakers, in the centre of the chamber, at

a distance of approximately 3 m. Helicopter noise from a BO-105 of the Norwegian Air Ambulance was used as a noise source. (Fig. 1). The method for recording, playback and frequency adjustment of this noise has been described in a previous paper (11).
Measurement and frequency analysis of the noise at the different altitudes used was subject to a separate pilot study.

Figure 1. BO-105 helicopter noise as used in the laboratory



In addition, noise levels were continuously monitored and logged throughout all experiments, using a Brüel & Kjær type 2135 noise level meter coupled to a personal computer with type 7636 statistical noise analysis software. The overall noise level for each altitude was thereby measured in all experiments.

Audiometry

Pure-tone audiometry was performed for the frequencies 500, 1000, 2000 and 4000 Hz at each altitude, using a Madsen Midimate 330 Audiometer (Madsen Electronics, 20, Vesterlundsvej, DK-2730, Herlev, Denmark) coupled to a personal computer. A Peltor aviation headset type MT32H7F-22 (Peltor, Box 2341, S-331 02 Värnamo, Sweden), was used. This headset had been adapted for audiometry by rewiring and fitting new plugs for channel separation. The headset is a widely used aviation headset with good noise damping characteristics. Speech audiometry was performed using one-syllable words. The word material was standard Norwegian speech audiometry test material, prepared on digital audio tape (Rikshospitalet, Oslo, Norway). Speech audiometry was performed for 8 signal levels with 5 decibel steps, starting from a fully audible level with approximately 100% score, down to almost inaudible. At each level 20 words were presented. Thus, 160 words in total were presented at each altitude.

Other clinical measurements

Tympanometry was performed before and after all experiments to ensure good tubal function and adequate equalisation on both ears.
An intra-arterial catheter was placed in the radial artery of each subject by a cardiologist, for collection of arterial blood samples for Haemoglobin Oxygen Saturation. Pulse oxymetry was also used for monitoring purposes during the experiment, but is a less reliable method. All subjects tolerated the arterial catheter well, with no severe discomfort.

Experimental design

The experimental staff consisted of four people: The chamber and audiometry operators outside the chamber, a physician with the subject inside the chamber, and a laboratory technician performing the arterial blood gas analyses. We used 4 different simulated altitudes for this experiment: sea level, 10.000, 13.000 and 16.000 feet. The order in which the 4 different altitudes were used, was subjected to certain criteria: None of the experiments were to be started or ended with the maximum simulated altitude, and no experiment should involve 3 or more subsequent increases or decreases, respectively, in altitude. These criteria left us with 8 different orders which were randomised. The experimental procedure was double-blind, i.e. neither the audiometry operator nor the subject should know which altitude was being provided at any time. Therefore, only subjects who had no prior knowledge of altitude chamber operations were used. Also, the audiometry operator was placed out of view of chamber instrumentation, and no mention of altitude-related parameters were to be made by any of the experimental staff except in an emergency. The experimental protocol was not opened until all the experiments had been completed.

For each simulated altitude, we performed the following procedure:

1. Ensuring the subject had full subjective middle ear equalisation.

2. Pure-tone audiometry (Chamber ventilation system off for this short period).

3. Arterial blood sample for haemoglobin oxygen saturation.

4. Noise on , speech audiometry after 2 minutes of noise (Total duration of noise : 11 minutes).

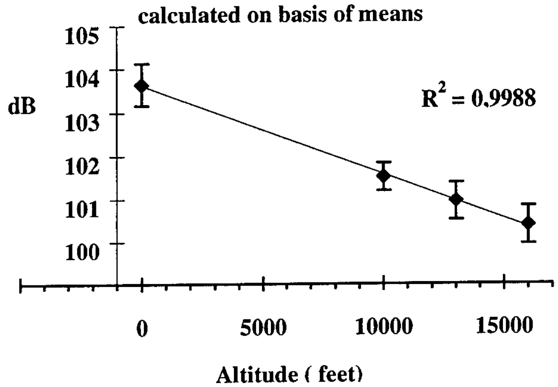
5. Noise off, arterial blood sample for Haemoglobin oxygen saturation.

The exact same procedure was followed for all the 4 simulated altitudes.
Evaluation of the above procedure during the experiments convinced us that the blinding of the experiments was successful, ruling out experimental bias both from the subjects and the audiometry operator.

Results

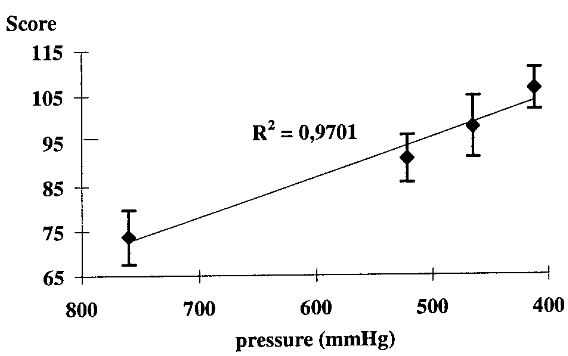
Noise levels gradually decreased with altitude (Fig. 2). This finding was expected, since it had been observed in the pilot study already mentioned.

Figure 2. Median noise level as function of altitude 95% confidence intervals shown. Linear regression calculated on basis of means



To verify that these findings were representative of an aviation environment, similar noise measurements were made in a RNoAF Twin Otter at different altitudes up to 15000 feet. The changes in noise level with altitude were much the same as in the altitude chamber.
Mean total speech intelligibility scores are shown in Figure 3, as a function of barometric pressure. These values are based on the total scores, i.e. the sum of the scores for all signal levels combined, for each altitude. Mean scores with 95% confidence intervals are shown.

Figure 3. Mean total speech intelligibility scores as a function of pressure with 95% confidence intervals shown. Regression line calculated on basis of means.



As one can see, there is a clear increase in speech intelligibility with altitude. The same data were also subjected to an analysis of variance (ANOVA). This statistical analysis shows a statistically highly significant difference between the mean scores for the different altitudes. An overview of ANOVA results is shown in table 1.

Table 1. ANOVA for total speech intelligibility scores at different altitudes.

Data	Mean	Variance	N
sea level	73.8	77.4	8
10,000 feet	91.1	49.8	8
13,000 feet	98.1	94.7	8
16,000 feet	106,5	44.2	8

Level of significance for difference of the means: p = 0.00000009

Speech intelligibility scores as a function of noise level is seen in Figure 4.
Likewise, speech intelligibility as a function of blood oxygen saturation is shown in Figure 5.

Figure 4. Mean total speech intelligibility scores as a function of median noise level, with 95% confidence intervals. Regression line calculated on basis of means.

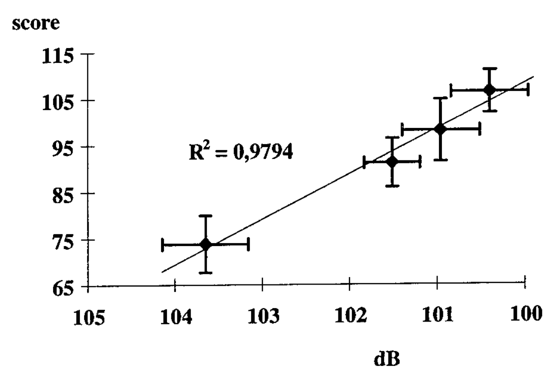
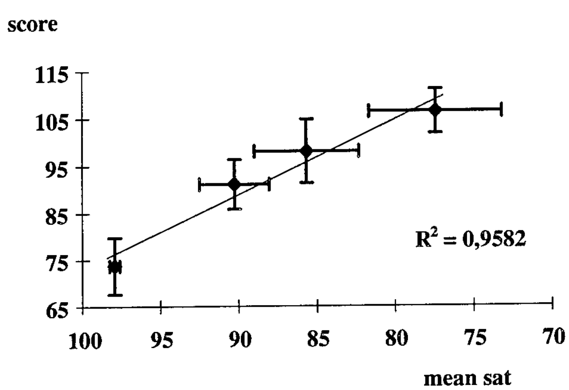


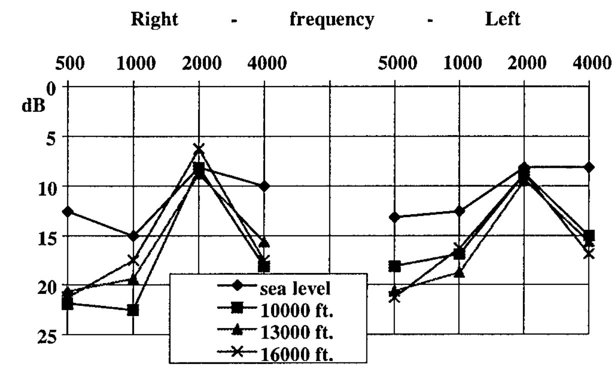
Figure 5. Mean total speech intelligibility score as a function of mean Haemoglobin oxygen saturation - Regression line calculated on basis of means. 95% confidence intervals for means are also shown. N=8



As is evident from these two figures, much the same regression is demonstrated. As the different parameters that change with altitude (noise, pressure, blood oxygen content) are so closely linked, results from e.g. a multiple regression analysis would not be helpful in separating out various factors.

Figure 6. shows the pure-tone audiometry results in this study. Mean thresholds are provided for the 4 simulated altitudes. As one can see, threshold changes with altitude are not uniform as the depend on the frequency. Thus, the 2000 Hz frequency seems to be relatively stable compared to the other frequencies measured.

Figure 6 Mean pure-tone hearing thresholds in dB for right and left ears, at the 4 different simulated altitudes



The statistical analysis on these data was performed by calculating the mean of the two ears (right and left), then performing an ANOVA for each frequency, comparing the means for different simulated altitudes. Thus we performed 4 ANOVA analyses, one for each frequency. A summary of the results are shown in table 2.

Table 2. ANOVA for difference of the means for the 4 simulated altitudes. Figures given for each frequency.

Frequency	Level of significance for difference of the means for the 4 simulated altitudes (p-value)	N
500 Hz	0.09	8
1000 Hz	0.22	8
2000 Hz	0.89	8
4000 Hz	0.02	8

Only the 4000 Hz frequency shows a significant change with altitude at an 0.05 level of significance.

Discussion

The results from the present set of experiments indicates a quite clear increase in speech intelligibility with altitude in aircraft noise. As this experimental setting has not been used earlier, there are no results to compare the present results with. However, this experiment should have a higher validity in relation to the aviation operational environment than any earlier report in this area.

The effect of hypoxia has previously mostly been the main subject in this area. However, any such effect is greatly outweighed by other effects, although Fig. 5 might indicate some drop-off in the trend for increased scores with altitude for the lowest haemoglobin oxygen saturation levels.

It appears that the reason for the increased speech intelligibility scores with altitude is related to noise level. Observing Figs. 2 and 6, it seems obvious that the decreased noise level with altitude is not paralleled by a similar decrease in auditory sensitivity. Earlier work already mentioned supports this aspect of our findings, especially relating to the different altitude-induced changes in threshold depending on frequency (4). Similarly, some work on auditory sensitivity in *hyper*baric air support that changes in auditory sensitivity are frequency-dependant (12-13), although it is not certain that we can extrapolate directly in this respect.

The results herein give a clear indication of an improved speech intelligibility with altitude in a setting mimicking an operational environment. The same results, however, raise some definite questions as to the nature of the mechanisms involved in the described findings:

1. How does decreased atmospheric pressure affect sound transmission inside all parts of the human ear?

2. How are signal-to-noise ratios for different frequencies affected by changes in pressure?

3. How can one correctly calibrate headsets and establish equivalent noise levels at different atmospheric pressures?

The last point, especially relating to headsets, is of course a reason for a certain degree of caution when judging the validity of these results in relation to other noise environments, protection equipment and communication systems.

Conclusions

The conclusions to be drawn from the present experimental findings may be summarised as follows:

1. Speech intelligibility in aircraft noise may show a substantial improvement with altitude

2. The reason for this improvement is probably related to an improved signal-to noise ratio brought on by

the change in atmospheric pressure. Hypoxia had little relative importance in our experimental setting.

3. Further research is warranted in order to clarify the underlying physical and physiological factors involved in the described changes.

Acknowledgements

This work was conducted with support from Norwegian Air Ambulance. The author wishes to thank Ole Tvete, Bjørn Ludvigsen and Harald T. Andersen for help in the planning and evaluation of the experiments. Thanks also to Carl Christian Christensen, Morten Ryg, Siw Robertsen, Toril Rode and Inger-Lise Neslein for help in performing the experiments. The author also wishes to thank 719 sqn, RNoAF, for assistance in the aircraft noise measurements at different altitudes.

References

1. Bronkhurst AW, Veltman JA, van Breda L. Application of a Three-Dimentional Auditory Display in a Flight Task. Hum Factors 1996;38(1):23-33.

2. Smith GM, Seitz CP. Speech Intelligibility Under Various Degrees of Anoxia. J.Appl.Psychol. 1946;30:182-91.

3. Smith GM. The Effect of Prolonged Mild Anoxia on Speech Intelligibility. J.Appl.Psychol. 1946;30:255-64.

4. Klein S, Mendelson S, Gallagher TJ. The Effects of Reduced Oxygen Intake on Auditory Threshold Shifts in a Quiet Environment. J.Comp.Physiol.Psychol. 1961;54:401-4.

5. Klein S. Effects of Reduced Oxygen Intake on Bone Conducted Hearing Thresholds in a Noisy Environment. Percept Mot Skills 1961;13:43-7.

6. Curry ET, Boys F. Effects of Oxygen on Hearing Acuity at Simulated Altitude. Eye, Ear, Nose & Throat Monthly 1956;35(April):239-45.

7. Burkett PR, Perrin WF. Hypoxia and Auditory Thresholds. Aviat. Space Environ. Med. 1976;47(6):649-51.

8. Fowler B, Lindeis A-E. The Effects of Hypoxia on Auditory Reaction Time and P300 Latency. Aviat. Space Environ. Med. 1992;63(November):976-81.

9. Fowler B, Prlic H. A Comparison of Visual and Auditory Reaction Time and P300 Latency Thresholds to Acute Hypoxia. Aviat Space Environ Med 1995 Jul;66(7):645-50.

10. Kitahara M, Ozawa H, Kodama A, Izukura H, Inoue S, Uchida K. Effect of Atmospheric Pressure on Hearing in Normal Subjects. Acta Otolaryngol Suppl (Stockh) 1994(510):87-91.

11. Wagstaff A, Tvete O, Ludvigsen B. The Effect of a Headset Leakage on Speech Intelligibility in Helicopter Noise. Aviat Space Environ Med 1996;67:1034-1038.

12. Thomas WG, Summit J, Farmer JC. Human auditory thresholds during deep, saturation helium-oxygen dives. J Acoust Soc Am 1974;55(4).

13. Fluur E, Adolfson J. Hearing in Hyperbaric Air. Aerospace Medicine 1966;37(August):783-85.

VULNERABILITY OF FEMALE SPEECH PRODUCED IN OPERATIONAL NOISES

C. W. Nixon, L. J. Morris, A. R. McCavitt, M. P. McDaniel
T. R. Anderson, R. L. McKinley and D. G. Yeager

Bioacoustics and Biocommunications Branch
Armstrong Laboratory, AL/CFBA
2610 Seventh Street
Wright-Patterson Air Force Base, Ohio 45433-7901, USA

1. SUMMARY

This study quantified the speech intelligibility differences in high level noise due to gender. Female speech was always less intelligible than male speech and the differences grew with increased levels of the noise. Intelligibility of both male and female speech differed with aircraft noise spectrum. These gender differences caused no impact at the lower levels of noise, however they do constitute a problem at the highest levels. The application of active noise reduction technology and replacement of the M-87 with the M-169 noise canceling microphone should neutralize most of these impacts. The perception of LPC-10 and CVSD vocoded female speech was essentially the same as male speech. There were no significant differences between the recognition accuracy of male and female speech for either the ITT or IBM automatic speech recognition system.

2. INTRODUCTION

Women are flying high performance aircraft and their increasing presence in the cockpits and crew stations of military strategic and tactical aircraft is assured. The design of current aircraft audio communication systems and components were optimized for male voice characteristics and may not fully accommodate the female voice. Current knowledge of the perception of female speech, particularly in the harsh environments of military aviation, is not sufficient to allow reliable estimates of female speech performance in the cockpit environment. This project studied the information necessary to identify significant differences, if present, in the perception of female and male speech. Differences that would prevent female speech from communicating effectively in current aircraft types were addressed. Difficulties with the perception of female speech would affect all aviators.

This research examined the perception of female speech produced in operational noise environments by listeners in operational noise environments. Emphasis was on female aviators and selected military aircraft noise environments that are typically present during use of military aircraft voice communication systems. Speech performance was measured in the cockpit noise environments of four different types of aircraft, with noise-canceling microphones, with digital speech coders and decoders, and with automatic speech recognition systems (voice controllers). Perception of female speech was evaluated relative to that of male speech and to performance criteria that indicate the effectiveness of speech communications under operational conditions.

3. RESEARCH OBJECTIVES

The research objectives of this study were to quantify the differences between the

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

intelligibility of female and male speech relative to those factors in military aircraft cockpits that influence voice communications, to determine whether the reductions in speech performance are or are not significant relative to operational noise environments, and to propose actions to minimize significant adverse effects, where feasible.

This research measured the extent to which female (and male) speech was affected by: (a) different cockpit noise environments (spectra) of four military aircraft, (b) the responses of standard military noise-canceling microphones, (c) digital encoding and decoding of the speech signals with the DoD standard LPC-10 (1) and Continuously Variable Slope Delta (CVSD) vocoders, and (d) ITT and IBM voice control or automatic speech recognition systems (2).

4. APPROACH

A four-phase study evaluated voice communication performance in a reasonable representation of operational conditions and speech communication technologies. The different aircraft noise spectra were selected to represent the range of cockpit noise environments in which female aviators are found. These included the low frequency spectra of the C-130E aircraft and MH-53 helicopter, the relatively flat spectrum (up to 4000 Hz) of the C-141B aircraft, and the higher frequency spectrum of the F-15A high performance fighter aircraft. The noise spectra are shown in Figure 1. In Phase I, speech performance was measured for each of the aircraft in the four different levels of the cockpit noise spectra. In Phase II, the relative effectiveness of the current standard noise-canceling microphones was examined in the same noise environments employed in Phase I.

The intelligibility of male and female speech processed by the DoD standard LPC-10 speech coder and a high quality speech coder (Continuously Variable Slope Delta modulation system, CVSD) was examined in Phase III. As noted earlier, the coder converts the analog speech signal to a digital signal that is transmitted to the receiver where it is reconverted to speech. Some fidelity of the speech signal is lost in this conversion process. Phase III examined the robustness of the reconstructed female speech in the presence of the four aircraft noise conditions of Phase I.

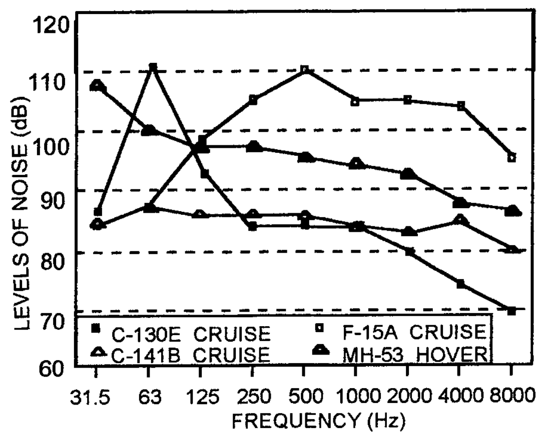


Figure 1. Spectra and Levels of Cockpit Noises of Aircraft During Cruise and Helicopter During Hover

In Phase IV, the recognition accuracy of female and male produced speech by two different automatic speech recognition (ASR) systems was evaluated in two cockpit noise environments. Recognition accuracy by ASR systems of male and female speech in high levels of aircraft cockpit noise had not been previously reported.

5. CRITERION MEASURE

The criterion measure of communication performance for Phases I, II, and III was the percent correct intelligibility of the Modified

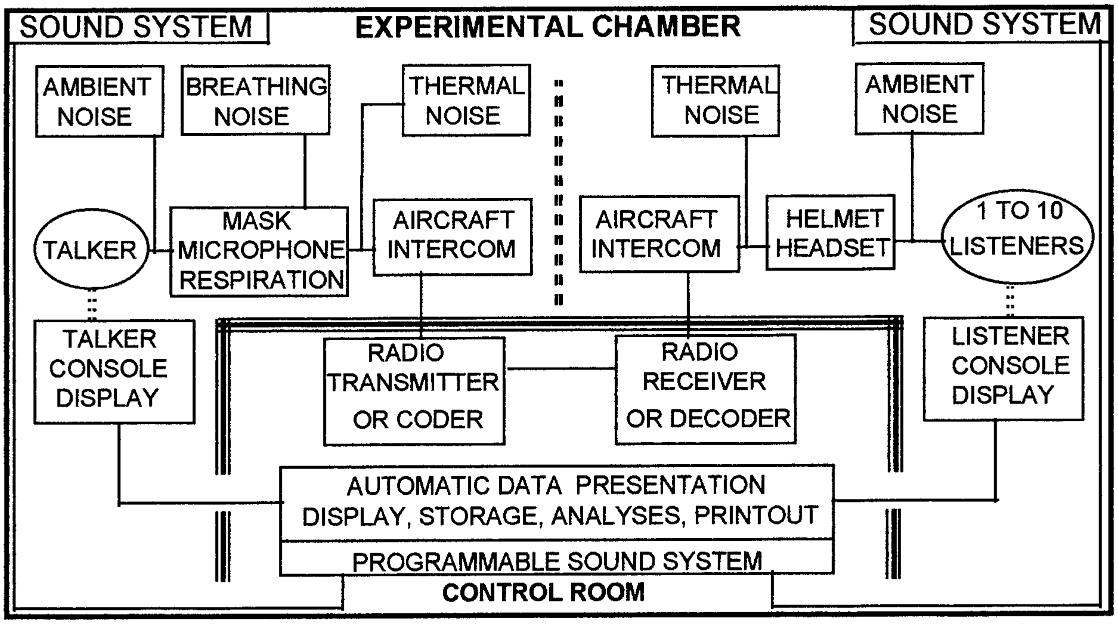


Figure 2. Configuration of the Voice Communications Research and Evaluation Facility Showing the Talker's Station and One of the Ten Listening Stations

Rhyme Test (MRT). The measurement of speech intelligibility was accomplished in accordance with the American National Standard, S3.2-1989, Method for Measuring the Intelligibility of Speech Over Communication Systems (3).

6. PERFORMANCE CRITERIA

The Armstrong Laboratory voice communications research facilities provide laboratory data that accurately represent speech intelligibility performance in corresponding operational situations. Criteria developed from these databases demonstrate that average laboratory speech intelligibility scores below about 70 percent correct are typically unacceptable in corresponding operational situations. Performance in the range from about 70 percent to 80 percent is marginal. Laboratory intelligibility of about 80 percent and above is acceptable under operational conditions. These "performance criteria" have been validated ***only*** for evaluations accomplished within the facilities and using the procedures of the Armstrong Laboratory.

7. SUBJECTS

All subjects were highly experienced communicators familiar with military communictions systems and with the use of helmets, headsets, oxygen masks and boom microphones in high levels of noise. They were recruited from the general civilian population and were paid an hourly wage for their participation. All spoke midwestern American English that was free from strong regional dialects and speech problems. Subjects exhibited normal hearing and middle ear function measured prior to participation in the study. Twenty adult subjects, ten male and ten female, participated as talkers and a subset of ten comprised of five males and five females, formed the listening panel. Sound attenuation of each headset/helmet worn by the subject was measured to confirm the individual hearing protection during the study (2).

8. FACILITIES AND EQUIPMENT

Data were collected in the Voice Communications Research and Evaluation System (VOCRES) housed in a large reverberation chamber (5). The configuration of the VOCRES architecture with a talker and a single listener station is shown in Figure 2. VOCRES contains ten individual voice communications stations that provide simultaneous communications and measurement of all test subject responses. Each station includes an alphanumeric light emitting diode (LED) display and a subject response unit consisting of special keyboards for entering response data. The stations also contain standard air respiration systems, aircraft voice intercommunications systems, and standard headsets/helmets for each aircraft noise environment as shown in Table 1. VOCRES also contains a programmable high intensity sound system that can duplicate the spectrum and level of any Air Force aviation noise environment.

AIRCRAFT NOISE	HEADSET/HELMET SYSTEM	MICROPHONE
C-130E	H-157 HEADSET	M-87
C-141B	H-157 HEADSET	M-87
F-15A	HGU-55/P HELMET	M-169 MBU/12P OXYGEN MASK
MH-53	SPH-4AF HELMET	M-87

Table 1. Headsets/Helmet Systems and Microphones Used with Corresponding Aircraft

9. PROCEDURES

Prior to data collection subjects were familiarized with talking and listening in the noise conditions. Listeners individually adjusted the gain of their stations to a comfortable listening level in each of the noises as is done in operational situations. The gain of the signals for talkers and listeners was not readjusted for the remainder of that condition.

During data collection, the listening panel members were seated at the voice communication stations in VOCRES and one of the twenty talkers was seated at the remote VOCRES station in the adjacent facility. Talkers and listeners were in identical noise environments during each experimental run. Subjects wore the individually fit custom-fit helmet or headset corresponding to the experimental condition being evaluated as shown in Table 1. For each experimental trial a test word appeared on the LED in front of the talker. The talker read the word in a carrier phrase. A list of six rhyming words appeared on the LED displays of the listeners; one was the spoken word. Each member of the listening panel selected the word she/he believed was spoken by pressing the response button adjacent to that word. After a pause of five seconds, the next word appeared on the LED of the talker and the procedure was repeated until the fifty words in the set were spoken and the subject responses collected. This constituted one trial. All data were compiled by the VOCRES central computer where both individual and average response scores were calculated. This procedure was followed for all experimental trials.

10. RESULTS

In Phase I and II, data were comprised of measurements of speech intelligibility of ten male and ten female talkers as perceived by a panel of ten listeners (five male and five female). In Phase III, data consisted of measurements of the intelligibility of coded and decoded speech of all talkers perceived by the panel of listeners. In Phase IV, data consisted of the word and sentence recognition accuracy of two speech recognition systems. The two hundred (twenty talkers $\times$ ten listeners) response scores were

averaged for each experimental condition. Means and standard deviations were calculated and differences among the means were evaluated using standard statistical paired t-tests at the 0.05 level of confidence.

Data were treated by measures of central tendency and variance with emphasis on the average differences between the means of the samples. The statistical significance of the differences between the means of the matched pairs (female and male) was determined by calculating the t-score and comparing it with the criterion t-value corresponding to the 95 percent confidence level (6) The statistically significant differences in many situations were so small they would be indistinguishable in the operational situation. The critical issue was whether the performance in a particular condition was acceptable (above 80), marginal (70 - 80), or unacceptable (below 70. These performance level criteria are indicated by dashed horizontal lines across figures where applicable.

11. PHASE I

Aircraft Cockpit Noise Spectra

The average intelligibility scores are summarized for the female and male subjects for the four aircraft and the ambient (66dB) and three levels of noise conditions. The data are shown in graphical form in Figure 3. The vertical bars on the figures represent plus and minus one standard deviation. Those differences between the female and male means that are statistically significant at the 95 percent level of confidence are circled on the graphs. The two dashed horizontal lines across figures indicate the boundaries of acceptable, marginal, and unacceptable performance.

The aircraft cockpit noise data in Figure 1 represent in-flight cruise conditions for which

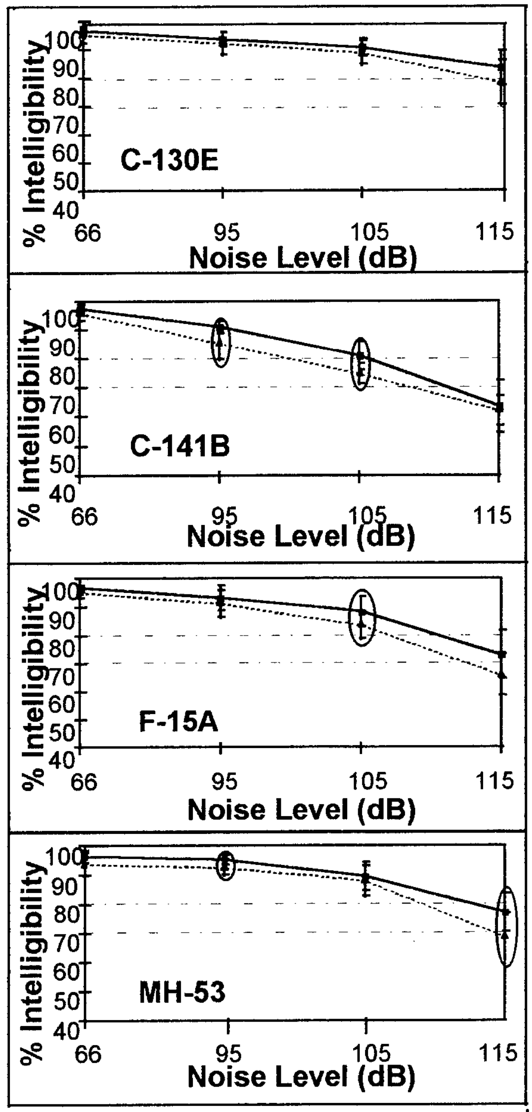


Figure 3. Summary Mean Intelligibility Scores for Female (.........) and Male (———) Talkers in the Aircraft Noise Spectra and Levels

the spectra and levels differ substantially among aircraft. In this study, the experimental conditions presented all the spectra at the same overall sound pressure levels (OASPL) shown in Figure 3. This was done to include the range of levels found in almost all operational aircraft, to allow comparisons among aircraft types, as well as to measure reductions in speech

performance as levels of noise spectra were increased for the individual aircraft.

The overall sound pressure levels of the noises measured during cruise and hover differed for each air vehicle ranging from a low of about 95 dB for the C-141B to a high of about 113 dB for the F-15A. The overall levels of about 110 dB of the C-130E and the MH-53 were determined by the higher levels of low frequency energy in their spectra. The overall sound pressure levels of the "cruise" spectra were adjusted to create the test condition levels of 95 dB, 105 dB, and 115 dB. The spectra for the 115 dB noise conditions for each aircraft type are shown in Figure 4. The C-141B cruise spectrum that was increased by 15 dB to reach the 115 dB level, approached the levels of the F-15A in the speech frequency region and exceeded that of the C-130E by 15 dB and that of the MH-53 by 5 dB. As a consequence, the C-141B had the highest and the C-130E the lowest level of noise in the speech frequency region for headset listening. The speech performance of both males and females was best in the C-130E and poorest in the C-141B. Speech performance was acceptable in all measured conditions for the C-130E and was unacceptable for both male and female speech at the highest level of the C-141B noise.

Decreases in intelligibility due to increases in level of the noise were measured in all aircraft spectra. The amount of the decrease became progressively larger with increasing levels of noise. For the C-130E, the decrease in intelligibility was three percent less at 105 dB than at 95 dB, and seven to 10 percent less at 115 than at 105 dB. The C-141B intelligibility was 10 to 11 percent less at 105 dB than at 95 dB and 13 to 17 percent lower at 115 dB than at 105 dB. These decreases in intelligibility were approximately the same for male and female speech except at the 115 dB, MH-53 condition where the decrease for females was larger and the 115 dB, C-141B where it was smaller.

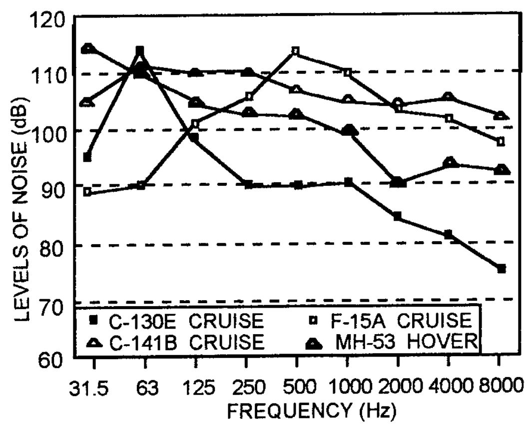


Figure 4. Noise Spectra and Levels for the Aircraft During the 115 dB Level of Noise Experimental Condition

C-130E Aircraft

Perception of the male and female speech was essentially the same at the 105 dB level of the C-130E noise and below with only a 5 percent difference at 115 dB as shown in Figure 3. None of the differences were statistically significant. Both male and female speech were around the 90 percent correct region and above at noise levels of 105 dB and below. At 115 dB, the accuracy's were 79 percent correct for females and 84 percent correct for the males; both were acceptable. The overall level of the noise of the C-130E during maximum endurance cruise was about 111 dB in the flight crew compartment and a maximum level of 115 dB at one of the other crew stations (5). Voice communication conditions in this aircraft, for female and male talkers, were considered acceptable.

C-141B Aircraft

The mean speech intelligibility of both males and females dropped almost 40 percent from the ambient to the 115 dB noise condition as shown

in Figure 3. The mean differences between genders at both the 95 dB and 105 dB noise conditions were statistically significant at the 95 percent confidence level. Both female and male speech were acceptable at the 95 dB level; at 105 dB, male speech was acceptable and female speech was marginal; and both were unacceptable at the 115 dB level. Assuming that the intelligibility function shown by the graph is linear, the extrapolated percent correct intelligibility at 100 dB should be almost 80 percent for the female and acceptable; it should be higher at lower levels of noise. The overall level of the noise measured between the pilot and copilot on the C-141B was almost 96 dB during cruise with a worst-case condition of 117 dB during taxi with four engines at taxi power and 111 dB during climb to 3000 feet. Communications will be acceptable during cruise, but unacceptable during taxi and climb.

F-15A Aircraft

The only statistically significant difference between the mean values shown in figure 3 occurred at the 105 dB noise condition which was acceptable for both genders. At the 115 dB level of noise, the male speech was marginal and the female speech unacceptable. The overall sound pressure level of the F-15A cockpit noise during cruise was about 110 dB and during a high speed run it was about 115 dB. The data suggest that female speech perception is marginal to unacceptable in the high noise environments of these two flight conditions and that the male speech is in the marginal region at 115 dB. Improvement is required for female speech to be understood by other aviators in the 110 dB - 115 dB levels of noise.

MH-53 Helicopter

Statistically significant differences between male and female speech perception occurred at the 95 dB and 115 dB noise conditions as shown in Figure 3. The small difference of only about 2.5 percent at the 95 dB noise condition was statistically significant. The mean difference at the 115 dB level of noise was about 8 percent. The speech perception of both female and male was acceptable at all except the 115 dB condition. At 115 dB, male speech was in the marginal region, close to the acceptable range. The female speech was a little below the marginal region and must be considered unacceptable. Improvement in female speech perception is required in these high level noises for good recognition by other aviators.

12. PHASE II

The conditions in Phase I in which the M-87 noise-canceling microphone was used were repeated in Phase II with the M-162 noise-canceling microphone. These two sets of data (Phase I M-87 microphone and Phase II M-162 microphone) were compared to evaluate the relative effectiveness in noise of the microphones with female and male produced speech. The M-169 oxygen mask noise-canceling microphone was not included in this evaluation. Since no alternative mask microphone is available, the M-169 data collected in Phase I represent its performance in the spectra and levels of the noises of interest.

The mean female and male speech intelligibility for the M-87 and M-162 noise-canceling microphones in the various levels of the aircraft spectra is shown in Figure 5. No statistically significant differences between female and male speech were observed with the M-162 microphone. All performance was acceptable, according to the performance criteria, except for the 115 dB noise condition for the C-141B aircraft which was unacceptable for both male and female talkers.

Mean speech intelligibility with the M-162 was better than with the M-87 for all aircraft and all

levels of noise. Female and male speech perception with the M-162 was acceptable in all conditions except the C-141B at the 115 dB level of noise. Female speech performance with the M-87 was marginal, and with the M-162 was acceptable in the C-130E at the 115 dB level of noise. Both microphones are unacceptable for the C-141B 115 dB noise condition and the M-87 was marginal in the MH-53 helicopter spectrum at 115 dB of noise.

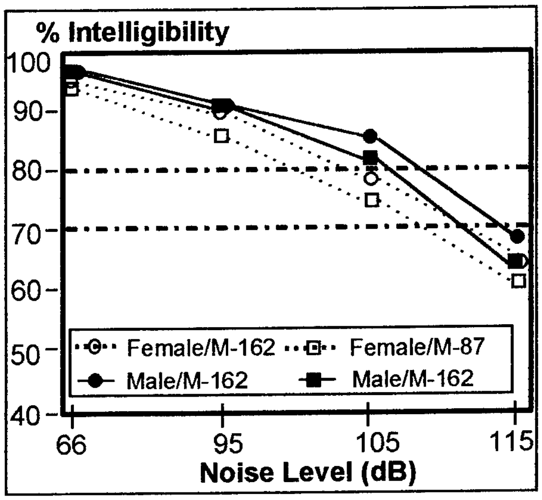


Figure 5. Mean Speech Intelligibility of Male and Female Talkers with the M-87 and M-162 Noise Canceling Microphones in C-141B Aircraft Cockpit Noise

The Phase II data indicate that the mean female speech perception was lower than the mean male speech perception for both microphones in all conditions; however, the amount of difference was relatively small and not statistically significant. These data suggest that the perception of both female and male speech may be improved in the three aircraft cockpits at all noise conditions by replacing the M-87 microphone with the M-162 microphone.

13. PHASE III

In Phase III, a remote talker station was located in an adjacent voice communication research facility that contains identical features as in the VOCRES. The talker was seated at this remote communication station and the listeners remained at their individual stations in VOCRES. Data were collected for vocoded female and male speech in the four aircraft noise spectra at four levels of each of the noises.

Linear Predictive Coder

There were no statistically significant differences between the perception of the female and the male speech in the 66 dB and 95 dB noise conditions. All of the aircraft communications were acceptable in the ambient condition, ranging from 80 to 85 percent correct responses. All of the aircraft communications were marginal in the noises at the levels of 95 dB, ranging from 72 to 78 percent correct as shown in Figure 6.

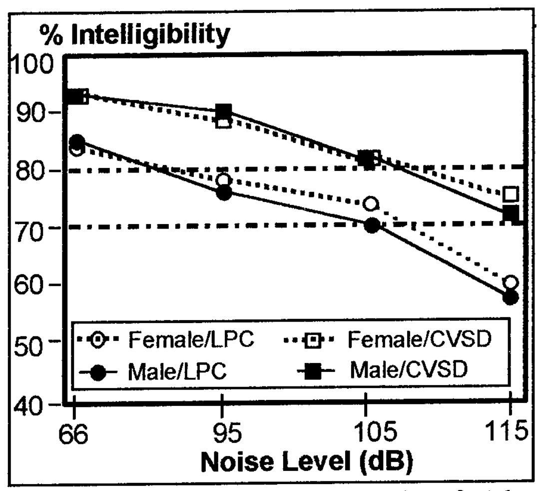


Figure 6. Mean Intelligibility in Noise of Male and Female Speech that was Coded-Decoded by LPC-10 and CVSD Vocoders

Communications were no better than marginal for all aircraft noise conditions at 105 dB, except the F-15A where female speech was unacceptable, and the C-141B where the speech of both genders was unacceptable. Average values range from 59 to 74 percent correct.

Voice communications were unacceptable for all conditions in the noises at 115 dB, ranging from 46 to 67 percent correct, except for the male speech that was marginal in the MH-53 noise.

Overall, the perception of LPC-10 coded female and male speech was acceptable in the ambient condition, marginal in the 95 dB level, and unacceptable in the 105 and 115 dB levels of the noises. Perception of the female and male speech was very similar in the lower levels of the noise. At the higher levels of noise, the female speech tended to be a little less intelligible than the male speech. Lesser intelligibility was statistically significant at only the two conditions cited earlier.

Continuously Variable Slope Delta Coder (CVSD)

There were no statistically significant differences between the perception of the female and male speech for the 66 dB and 95 dB conditions. Generally, these voice communications were acceptable with values ranging from about 77 to 94 percent correct as shown in Figure 6. The only statistically significant differences between the female and male speech were with the F-15A noise at levels of 105 and 115 dB. At 105 dB, male speech was acceptable and female speech was marginal; at noise levels of 115 dB, male speech was marginal and female speech was unacceptable. The perception of female and male speech was virtually the same in the other three aircraft noises at all four levels. Both female and male speech were unacceptable for all aircraft noises at 115 dB, except for the C-130E where both were marginal. As with most other factors examined in different levels of noise, perception of female speech tended to decrease more than male speech as the levels of the noises increased to the highest measured levels.

LPC-10 and CVSD Performance

The perception of the female speech is almost the same as the male speech when both are processed by either one or the other vocoder. Only four of the thirty-two measurement conditions showed statistically significant differences due to gender in percent correct intelligibility. Overall, the perception of the female speech was equally as effective as the male speech for either vocoder.

Although the intelligibility of the female and male speech was very similar for either vocoder, the differences between the performance of the two vocoders were statistically significant at almost all conditions. In all test conditions, the average percent correct intelligibility of the CVSD speech was higher than the LPC-10 data. Differences between the vocoder mean values as a function of level of the noises were as high as 15 percent. Sixteen of the conditions with the CVSD and only six with the LPC-10 were acceptable while six of the CVSD and ten of the LPC-10 were unacceptable, based on the performance criteria.

14. PHASE IV

AUTOMATIC SPEECH RECOGNITION (VOICE CONTROL)

Recognition accuracy of a speech recognition system is generally measured at two levels: at the sentence level and the word level. The percent correct sentences and words are calculated as:

$$\%\ \text{Correct} = \frac{\text{Number Correct}}{\text{Total Number}} \times 100\%$$

The Phase IV criterion measure is recognition accuracy calculated using this formula.

ITT VRS-1290 Speaker-Dependent ASR System

There were no significant differences between the recognition accuracy of the female and of male speech in any of the sixteen experimental conditions. Female and male produced speech were recognized with equal accuracy. Single word recognition was 90 to 97 percent correct and sentence recognition was 10 to 15 percent less (66 to 86 percent) in the C-130E noises. The ITT ASR system word and sentence recognition were resistant to the increased levels of the noise showing a relatively flat response with a slight reduction only at the 115 dB noise condition.

Both word and sentence recognition were very vulnerable to the MH-53 noise spectra and the increased levels. Word recognition dropped from an acceptable 88 percent correct in the 95 dB noise to an unacceptable 40 percent in the 115 dB noise condition. Sentence recognition was unacceptable at all three noise levels with about 60 percent correct in the 95 dB noise level and dropping to less than 10 percent in the 115 dB level of noise.

Overall, the ITT ASR system worked very well in the C-130E noise spectrum and levels. However, the same system operated at an unacceptable level in the noise spectrum of the MH-53 helicopter. The ITT ASR system was robust in increasing levels of the C-130E noise spectrum and should be acceptable under operational conditions. The ITT ASR word recognition was unacceptable for all MH-53 noise conditions of about 105 dB and above. Sentence recognition accuracy in the MH-53 spectrum was unacceptable in all noise conditions of 95 dB and above. These data indicate that those working on ASR applications in military aircraft must accomplish more work on ITT ASR performance as a function of spectrum in order to achieve greater operational effectiveness in noise spectra similar to that of the MH-53 helicopter.

IBM VoiceType ASR System

There were no statistically significant differences between the IBM ASR recognition accuracy of the female and of the male speech, with the exception of the sentence recognition in the MH-53 noise spectrum at 115 dB. This exception is attributed, in part, to the greater improvement in performance of the male speech at 115 dB over that of the 105 dB data. Single word recognition was in the marginal range of 70 to 83 percent correct and sentence recognition was only 48 to 65 percent (95 dB condition) correct in the C-130E noise spectrum. Recognition accuracy was relatively resistant to the increased levels of both the C-130E and MH-53 noise spectra. A maximum reduction of only 14 percent was observed for the word recognition accuracy and 13 percent for the sentence recognition accuracy over the approximate 50 dB increase in levels (66 dB to 115 dB) of the C-130E noise. Corresponding reductions in the MH-53 noises were 26 percent for the word recognition and 34 percent for the sentence recognition. The standard deviations of the male speech were generally larger than those of the female speech when using the IBM ASR system.

Overall, the IBM system exhibited some resistance to degradation by high levels of the two aircraft noises used in this phase of the study, and showed lowest scores of around 35 and 40 percent correct. However, it also shows some limitations by its inability to perform at higher than 85 percent correct in the low level ambient noise condition, which had a relatively flat spectrum and a moderate level of 66 dB.

The IBM system operated similarly in both spectra, suggesting its potential for a broader range of applications.

15. SUMMARY

Overall results from Phases I, II, and III revealed that the mean percent correct intelligibility of female produced speech was lower than the mean intelligibility of male speech. The amount of the difference increased as the level of the noise condition increased. The maximum effect usually occurred at the condition of highest level of noise.

Phase III data indicated that female speech was not significantly more vulnerable than male speech to specific vocoders. However, perception of female speech using the DoD standard LPC-10 vocoder was unacceptable in all four aircraft noises at the levels of 105 and 115 dB. Difficulties may be expected in the operational situation for both males and females at these levels.

It is noted that the IBM ASR female speech recognition scores were higher than the male scores for all ambient conditions and for the 95 dB level of the C-130E noise spectrum. All other conditions displayed the same values for both male and female or the usually observed slightly lower values for the female speech recognition.

There was no significant difference between the word or sentence speech recognition accuracy of male and female speech in the cockpit noise conditions evaluated by wither the ITT or IBM ASR. However, it is clear that additional work needs to be accomplished to improve overall performance of all ASR systems in cockpits and all other high level noise conditions.

16. RECOMMENDATIONS

Interpretations of the data suggest that the following actions might alleviate most of the noise induced gender related voice communications deficiencies identified: (1) Replace the standard M-87 noise-canceling microphones with the M-162 noise-canceling microphones; (2) Provide headsets and helmets with appropriate active noise reduction (ANR) technology capability; (3) and Complete development of a lightweight ANR headset for non-flight helmet applications such as C-130E and C-141 type aircraft. Also, (4) Upgrade LPC-10 vocoder algorithms and provide a good noise exclusion headset for the listener and an effective microphone noise shield for the talker, and (5) ASR systems should be evaluated in the environment (i.e., noise, vibration, heat) in which they will be used prior to their acquisition and installation. Additional work is required to improve overall performance of ASR systems in high level noise conditions.

17. REFERENCES

1. Nixon, C. W. and McKinley, R. L., "LPC-10 Intelligibility of Oxygen Masks and Microphones in Noise," Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB OH 45433, AAMRL-TR-88-048, November 1988.

2. Simpson, Carol, "Evaluation of Speech Recognizers for use in Advanced Combat Helicopter Crew Station Research and Development," Ames Research Center, NASA Contractor Report 177547, US Army Aviation Systems Command, TM-90-A-001, Moffet Field CA 94035, March 1990.

3. ANSI S3.2-1989 (ASA 85), American National Standard Method for Measuring the Intelligibility of Speech Over Communications Systems.

4. ANSI S12.6-1984 (R 1990), American National Standard Method for the Measurement of Real-Ear Attenuation of Hearing Protectors.

5. McKinley, Richard L., "Voice Communication Research and Evaluation System," Aerospace Medical Research Laboratory, AMRL-TR-80-25, Wright-Patterson AFB OH, 1980.

6. Box, George E. P., Hunter, William G., and Hunter, J. Stuart, Statistics for Experimenters, John Wiley and Sons, Inc., 1978.

A VOICE COMMUNICATION EFFECTIVENESS TEST

Bioacoustics and Biocommunications Branch
Armstrong Laboratory, AL/CFBA
2610 Seventh Street
Wright-Patterson Air Force Base, Ohio 45433-7901, USA

Richard L. McKinley
Thomas J. Moore
Charles W. Nixon

1. SUMMARY

A Voice Communication Effectiveness Test (VCET) was developed for relating voice communications performance to the effective completion of tasks with varying complexity, criticality, and time constraints. VCET is based on information theory and was initially developed for military applications. This metric accounts for the information required for task completion, the time available for the task, and the criticality or cost of an error relating to the quality of the communication channel. It can quantify the effects of competing workload on voice communications and it encompasses a wide range of voice communications systems and equipment, noise environments, and military missions. The rationale, development, and performance of these powerful analytical tools with applications in both the military and civilian communities were described. VCET shows great promise as the first true voice communication effectiveness test measuring not only intelligibility, but also information transfer with or without time dependency.

2. INTRODUCTION

Voice communications are vital to the successful execution of the simplest to the most complex military operation. Efforts have been focused on maintaining a high level of information transfer under all operational conditions. Comparatively small numbers of efforts have been accomplished to investigate components of the communication environments that relate performance levels of non-communication tasks to percent correct speech intelligibility. Few of these have been verified outside the laboratory. Even less attention has been directed to other elements relating voice communications to successful task completion, time required to accomplish tasks, and the simplicity or complexity of the task.

Voice communication is the primary mode of human information transfer for the accomplishment of numerous types of tasks. Voice communications are endemic to almost all facets of human interaction. As technology reduces the number of repetitive tasks to be accomplished by humans, information transfer becomes increasingly important. These facts make understanding voice communication and information transfer even more important than in the past. Military and civilian designers, builders, and users of voice communication systems require accurate and stable metrics to measure the performance of new voice communication hardware, software, vocabularies, syntaxes, etc., in the environmental setting of their expected use.

3. BACKGROUND

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

"Voice communication effectiveness" is defined as the transmission and reception of information required for the timely and successful accomplishment of a specific task. Attempts to measure voice communication effectiveness have been relatively few, whereas those that measure voice communication accuracy or intelligibility are many and varied. In general, the majority of these tests were developed as diagnostic tools for use in clinical settings to facilitate the treatment of problems with speech, hearing, and language. Many have been widely adopted and continue to be used in clinics today. Word and syllable type tests, and many comprised of sentences, are designed to measure recognition, i.e., the listener does not need to know the word but simply to recognize and identify the acoustic signal that was heard. The applications and types of these clinical tests expanded to other areas. They were modified or adapted, and others newly developed, for various applications involving person-to-person communications and systems commonly found in everyday life (telephones, public address systems, etc.).

For purposes of this discussion, these tests are separated into the three basic categories of clinical diagnostic tests, speech intelligibility tests, and predictive measures of speech intelligibility. Substantial overlap exists among these tests and the differences between them are frequently where they are applied. All of these, at one time or another, have been used to estimate various aspects of voice communication including accuracy. All of them measure parameters which may be related to or a portion of voice communication effectiveness but none seem to encompass voice communication effectiveness.

Clinical Diagnostic Tests

The diagnostic speech tests were developed for the audiologists for use in the evaluation of syllable, word, and phrase perception in the diagnosis and treatment of problems with speech, hearing, and language and with the fitting of hearing aids. Some randomly selected samples are the spondee words originally developed by Hudgins, et. al.(1), in 1947 and later modified by Hirsh et. al.(2), in 1952, the rhyme test developed by Fairbanks in 1958 (3), the speech perception in noise (SPIN) test developed by Kalikow et. al. (4) in 1977, and the hearing in noise test (HINT) developed by Soli in 1994 (5). Overall, the majority of the clinical speech tests are very good diagnostic tools while also providing some information on recognition and intelligibility performance. These successful clinical tests, as a class, do not provide the robust information required for a measure of communication effectiveness.

Speech Intelligibility Tests

Many researchers of voice communications have chosen a speech intelligibility test for use as a measure of voice communication performance. These intelligibility tests have taken many forms including, sentences, polysyllabic words, monosyllabic words, phonetically balanced words, and nonsense syllables. Fairbanks developed the original rhyme test, that was later modified and improved by House, et. al. (6), into the modified rhyme test (MRT) for use in the evaluation of voice communication systems. The diagnostic rhyme test (DRT), described by Voiers (7), has both diagnostic and intelligibility measuring capabilities. The DRT was designed to analyze speech communication systems and associated components. The acoustic characteristics of the test words that passed through the system and were perceived incorrectly are used to identify elements in the communication system

needing improvement. Many of these tests gained widespread acceptance by the scientific and technical communities. Three of them formed the basis of an American National Standards Institute (ANSI) standard procedure S3.2-1989, *Methods for Measuring the Speech Intelligibility of Communication Systems* (8). This standard describes both the speech test materials and the applications of those materials in the measurement of the performance of speech communication systems. However, none of these tests account for the variability in task complexity or time to complete the task as required for an acceptable measure of voice communication effectiveness.

Predictive Measures of Intelligibility

Two predictive measures of speech intelligibility have gained broad acceptance in the scientific community. These are the Articulation Index (AI, French and Steinberg, 1947)(9), as described by Kryter in 1962 (10,11) and again in 1969 in ANSI S3.5, and the Speech Transmission Index (STI), described by Houtgast and Steeneken in 1982 (12). The AI is a weighted speech-to-noise ratio with corrections for interfering parameters such as reverberation and peak clipping. The STI employs a modulated speech-like test signal that is measured relative to the noise level (the signal-to-noise ratio) in various frequency bands. A-weighted summation of these bands provides the STI numerical value. Both the AI and the STI provide a number between 0 and 1 that is correlated with speech intelligibility in linear communication environments. Many communication systems, however, are non-linear and both the AI and STI can predict speech intelligibility for these systems that is significantly different than speech intelligibility measured by panels of human listeners. These predictive methods have application in the design and development of communication equipment, but their inaccuracy in many practical communication environments makes the use of human listening panels the only valid method of determining speech intelligibility performance. The predictive methodologies do not provide a basis for a voice communication effectiveness measure.

Some researchers, such as Astrid-Schmidt Neilson (personal communication) (13), have attempted to address some of the elements of voice communication effectiveness in the development of communicability tests. These tests usually involve two-way communications that describe a picture or provide voice interactions while playing a game such as battleship. These two-way communications tests include the feedback loop in the communication system but do not address or control other factors in communication effectiveness such vocabulary size, phrase structure, and time. However, these interactive communicability tests are a step in the right direction.

No current metric accounts for voice communication effectiveness factors, such as vocabulary size, communication syntax, task complexity, the criticality of the task, and time to complete the task. These factors have created multiple requirements for speech intelligibility for different tasks or for times available to complete the tasks for task specific voice communications. None of the current speech evaluation procedures measure or are highly correlated with voice communication effectiveness across a wide range of conditions.

4. OBJECTIVE

The objective of this report is to describe the development and application of an information theory based procedure for the measurement

of voice communication effectiveness. Research was accomplished to implement the concept, design, fabrication, and verification of a voice communication effectiveness metric. The metric accounts for different intelligibility requirements, ranges of task complexity, task criticality, and time. The influence of environmental and voice communication link parameters in the voice communication environment are included in the process. The metric is reliable, stable, and has reasonable face validity.

5. CONCEPT

The concept for VCET is based on information theory. The hypothesis is that humans take advantage of the probabilities and statistics of voice communications to enhance communications effectiveness in settings which require task completion. Completion of tasks requires the transfer of information within a given time, and with minimum costs associated with errors. Analysis of the information content of suitable voice communication traffic provides statistics that define the range of the information values for the traffic. The data derived from the analysis forms the basic information requirements for a voice communication effectiveness test. Since information transfer frequently involves two-way interaction, VCET allows and accounts for such interaction.

6. APPROACH

Information analyses were accomplished on over 2,500 hours of military in-flight and ground based voice communications traffic. These analyses included voice communications intelligibility and communicability tests, and tasks that require communications for their completion. The information derived was employed to create an information theory based model of human voice communication effectiveness. It was readily apparent from this model that a tool was required to provide direct measurement of information transfer in a voice communication environment with or without competing tasks.

The approach was to map voice communication parameters to information theory terms so that the powerful analytical mathematical tools inherent in information theory could be used in the analysis tool. The critical information theory parameters to be measure or derived were entropy (a measure of randomness of the vocabulary), mutual information (a measure of amount of information that could be inferred about the next word to be received from the previous word received), channel rate (the amount of information attempting to be transmitted), and the channel capacity (the maximum amount of information which can be transferred over a given voice communication channel within the constraints of the entropy (vocabulary) and mutual information (message syntax).

7. INFORMATION ANALYSES

Information analyses were conducted on standardized speech intelligibility test materials and on an extensive voice communication audio tape data base. These materials were analyzed using the statistical mathematical tools from information theory (Shannon 1948 (14), Gallager 1968 (15)). Shannon developed these tools to deal with the basic aspects of communication systems. In Shannon's description of information theory is the concept of a "bit of information". A bit of information is defined as the amount of information gained as the result of the answer to a yes/no question. It is not a "bit" as in computers, but it can be when the vocabulary or alphabet is only two states such as "1" and "0". Shannon describes other terms such as entropy, mutual information, and channel

capacity. These terms are discussed later relative to their application to VCET, but for a thorough description see Shannon 1948 or Gallager 1968.

Standardized speech intelligibility test materials were analyzed for their information content in terms of entropy. Entropy can be thought of as a measure of the randomness or uncertainty of the vocabulary. The analysis assumed a uniform probability distribution across the vocabulary (i.e. all vocabulary items were equally probable). The effects on speech intelligibility of varying channel capacity, as measured by the weighted signal-to-noise ratio Articulation Index, are shown in Figure 1. The two upper curves are for 32 sentences or 32 individual words, the entropy of these materials is 5 bits. The middle two curves are for 256 phonetically balanced words and rhyme tests of 300 words with entropies of 8 bits and 8.2 bits respectively. The lower two curves are for vocabularies of 1,000 phonetically balanced words and 1,000 nonsense syllables with entropies of 9.96 bits. The effect of the uncertainty in the vocabulary can be seen in the speech intelligibility performance with varying weighted signal-to-noise ratios or AI. Additionally, two popular rhyme tests, the Modified Rhyme Test and Diagnostic Rhyme Test were analyzed. Their 300 word vocabularies give entropies of 8.2 bits. Mutual information analysis for the MRT and DRT gives 5.6 bits per phrase and 6.6 bits per phrase respectively. Once the response set is seen, the information transfer in the MRT is 2.6 bits and in the DRT is 1 bit.

Insert Figure 1 about here

8. VOICE COMMUNICATION DATABASE ANALYSES

Audio tapes of many types of voice communications conversations were obtained from the Air Force, Army, and Navy. A group of experienced listeners transcribed the tapes for the information analysis. The analysis measured the entropy of the vocabularies used for each type of task. The measurement of entropy involved identifying each of the vocabulary items used in communications for the given task and then counting the frequency of use of each of the vocabulary items. The frequency counts were used to compute probabilities for each of the vocabulary items. The higher the probability of a given word, the less information is conveyed by that word. An example is a vocabulary with only one word. No information is transferred using only that one word vocabulary; the receiver already knows what it is when it is transmitted. The probabilities of the vocabulary items were used in the computation of the entropy of the vocabulary as in equation 1.

$$H[I(a_y)] = \sum_{k=1}^{\infty} P(a_k)\log_2\left(\frac{1}{P(a_k)}\right)$$

Equation 1

As the probability of each vocabulary item approaches a uniform value across the vocabulary the entropy increases. Likewise, as the probability of one or a few items significantly increases relative to the rest of the vocabulary, the entropy decreases.

The second information analysis conducted on the transcribed message traffic was a computation of mutual information. Mutual information can be thought of as the amount of information derived about the next word to be received from the previous word received. Another way of thinking about mutual information in a voice communication context is that mutual information is the converse of the redundancy in the phrase or sentence. As

mutual information increases the amount of redundancy or predictability in the phrase or sentence decreases. As mutual information decreases the redundancy or predictability increases. Mutual information is calculated from the non-zero conditional probabilities of each of the vocabulary words paired with all other vocabulary words as in equation 2.

$$I(X;Y) = \sum_{k=1}^{K-1}\sum_{j=1}^{J-1} Q(k)P(j\_k)\log_2 \frac{P(j\_k)}{\sum_{i=0}^{K-1} Q(i)P(j\_i)}$$

Equation 2

The results of the mutual information analysis of the transcribed speech data that formed the basis for VCET are shown in Table 1. The 2,500 hours of taped speech were divided into 22 subgroups according to task. The vocabulary sizes for these various tasks range, an order of magnitude, from 204 words to 2089 words. The range of entropies for the same 22 subgroups is small with a minimum of 6.4 bits and a maximum of 7.5 bits. The mutual information range of 3.2 to 4.8 bits is also small. The application of the information analysis tools to this data set indicates good consistency across tasks. This uniformity allows the expansion of the applicability of this type of analysis and modeling to the voice communication area.

Insert Table 1 about here

Channel capacity is the maximum amount of information that can be transmitted over the communication system as in equation 3.

$$C \equiv \max_{Q(0),\ldots,Q(K-1)} \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} Q(k)P(j\_k)\log \frac{P(j\_k)}{\sum_{i=0}^{K-1} Q(i)P(j\_i)}$$

Equation 3

The theoretical channel capacity can be calculated by maximizing the mutual information over all the possible combinations of vocabulary items. This type of calculation in the context of VCET would give the channel capacity of the lexicon. The purpose of VCET, however, is to objectively measure channel capacity *in situ*. In order to accomplish this objective, additional analyses were performed on the speech data base to determine the statistical distribution of the information transmission rate or channel rate in bits per second and the amount of information per phrase in bits. The results from these analyses are shown in figures 2 and 3 respectively. The modal value for channel rate ranges from 9 to 15 bits per second while the modal value for bits per phrase ranges from 12 to 26 bits per phrase. These parameters of the information content and channel rate of actual communications were emulated in the development of the VCET vocabulary and phrase structure.

Insert Figures 2 - 3 about here

9. DESCRIPTION

The VCET vocabulary items were selected from the 2,500 hours of military voice communications audio tapes that were transcribed for the information analysis. In order to be used in VCET a vocabulary item needed to have at least three other words in the corpus that were rhyming words or words that were easily confused with the selected vocabulary item. In addition, the selected word was required to be a word contained in the non-zero conditional probabilities set, i.e., the word must be associated with other words in the vocabulary. The vocabulary items were then formed into two to six word phrases which met both the probability distribution functions of the original data set and met the syntactical rules of the original set. Samples

of the VCET test phrases are shown in Table 2. Table 3 details a list of confusable words that is used with the phrases cited in Table 2.

Insert Tables 2 and 3 about here

The amount of information that is in each of the phrases can be varied by modifying the syntactical rules of the phrases, i.e., varying the entropy at each of the nodes of the syntax. For example, a six word phrase could transfer as little as two bits of information or as much as 60 bits of information. This gives the researcher the ability to vary the channel rate (the information transmission rate) while maintaining the same vocabulary items. Therefore the channel rate can be varied to be above, equal to, or below the channel capacity of the communication link.

The basic VCET materials are two to six word phrases with each word having at least three other rhyme words. The task for the communicator(s) is to arrange the correct words in the phrase in the proper order in a minimum amount of time. It is a two-way, time dependent test. This structure of the test allows the sender and receiver to work together to maximize the information transfer rate. Word correctness, word order, and time are scored. Subject pairs with the best performance are rewarded with either a monetary reward ($0.25/hour bonus) or a visual reward (having their names posted outside the experimental chamber).

As the channel rate approaches the channel capacity the error rate should increase. The error rates from VCET, with known channel rates, give the user the capability to measure the channel capacity of the communication system under test. Application of this combined information allows the researcher to predict, within statistical confidence limits, the probabilities of successful completion of a range of complex tasks within defined time constraints.

The VCET components of vocabulary size, probability of individual vocabulary words, information requirements, and time to complete the voice communications task vary with the task or application.

10. RESULTS

VCET performance data has been obtained for a wide range of communication conditions. Performance data for one of these conditions also showing changes that occurred in various levels of noise is displayed in Figure 4.

11. COMMENT

VCET provides the expected type of results as shown in figures 2 and 3. The procedure is sensitive to changes in signal-to-noise ratio, bandwidth, distortion, clipping, and reverberation. VCET has not undergone rigorous validation. Overall, VCET shows great promise as the first true voice communication effectiveness test measuring not only intelligibility but also information transfer with or without time dependence.

SUMMARY

The results demonstrate the fundamental basis of VCET. It can be used in any communication application in the military in addition to most of the applications in the civilian environment. VCET is a stable and relatively sensitive measure of voice communication effectiveness which, when used with information based models of voice communications, can form a very effective measurement and/or analytical tool. This report described the background, structure, and preliminary performance of VCET. Future efforts will focus on the expansion of

VCET and rigorous field validation of the results in a wider range of tasks and environments.

REFERENCES

1. Hudgins, C.V., Hawkins, J. E., Karlin, J. E., and Stevens, S. S., " The Development of Recorded Auditory Tests for Measuring Hearing Loss for Speech," Laryngoscope, 57, 57-89, 1947.

2. Hirsh, I. J., Davis, H., Silverman, S. R., Reynolds, E. G., Elbert, E. and Benson, R. W., "Development of Materials for Speech Audiometry," Journal of Speech and Hearing Disorders, 16, 321-328, 1952.

3. Fairbanks, G., "Test of Phonemic Differentiation: The Rhyme Test," Journal of the Acoustical Society of America, 30, 596-600, 1958.

4. Kalilkow, D. N., Stevens, K. N. and Elliott, L. L., "Development of a Test of Speech Intelligibility in Noise Using Sentence Materials with Controlled Word Predictability," Journal of Acoustical Society of America, 61, 1337-1361, 1977.

5. Soli, Sigfrid, "Objective Measurement of Speech Intelligibility in Noise: The Hearing in Noise Test (HINT)," Proceedings of the National Hearing Conservation Association, Atlanta, GA, February 1994.

6. House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D., "Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set, " Journal of Acoustical Society of America, 37 (1965), 158-166.

7. Voiers, W. D., "Diagnostic Evaluation of Speech Intelligibility," In Speech Intelligibility and Speaker Recognition, M. E. Hawley, ed. Stroudsburg, PA, Dowden, Hutchinson & Ross, 374-387, 1977.

8. ANSI S3.2-1989 (ASA 85), American National Standard Method for Measuring Intelligibility of Speech over Communication systems

9. French, N. R. And Steinberg, J. C., "Factors Governing the Intelligibility of Speech Sounds," Journal of Acoustical Society of America, 19, 1947, 90-119.

10. Kryter, K. D., "Methods for the Calculation and Use of the Articulation Index," Journal of the Acoustical Society of America, 34, 1689-1697, 1962.

11. Kryter, K. D., "Validation of the Articulation Index," Journal of the Acoustical Society of America, 34, 1698-1702, 1962.

12. Houtgast, T. and Steeneken, H. J. M., "STI - An objective Measure for The Performance of Voice Communication Systems," AGARD-CP-3111, NATO, France, 17-1 to 17-6, March-April 1981.

13. Astrid-Schmidt Nielson, Naval Research Laboratory, Code 5532, Washington DC 20375-5337, (Personal Communication).

14. Shannon, C. E., "A Mathematical Theory of Communication," University Press, Urbana, 1949.

15. Gallager, R. G., "Information Theory," in The Mathematics of Physics and Chemistry, H. Margenau & G. M. Murphy (eds.), Vol 2, Van Nostrand, Princeton, NJ, 1964.

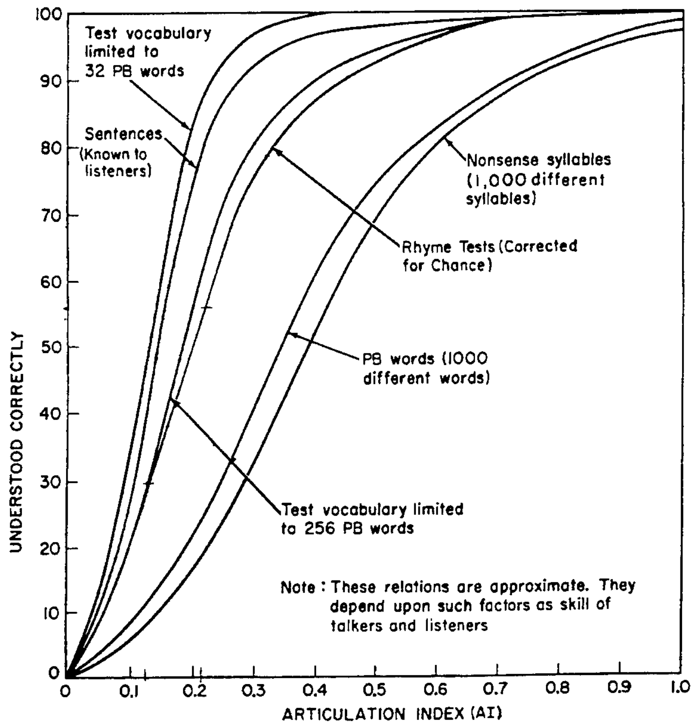


FIGURE 1.

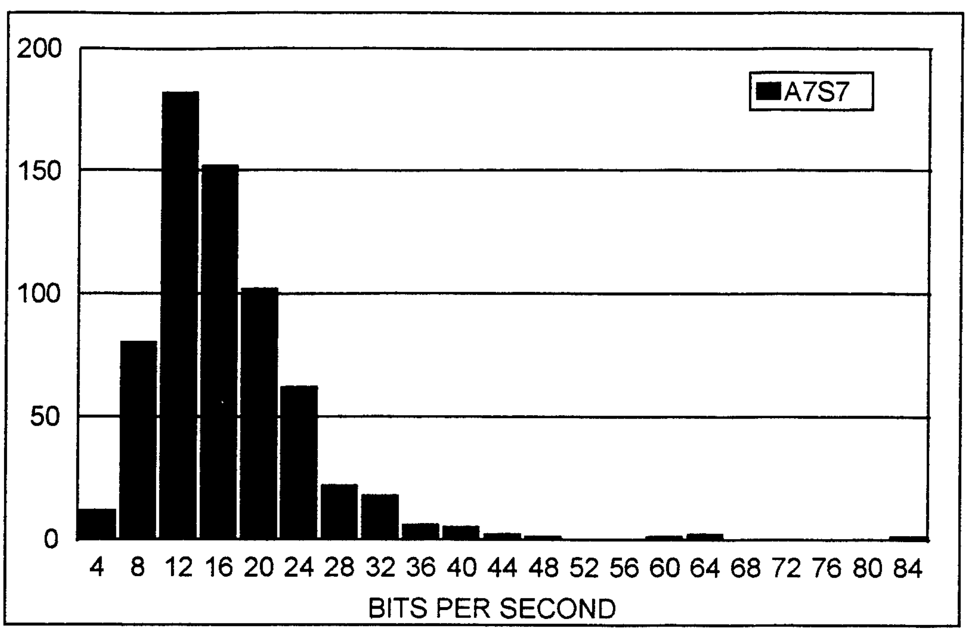


Figure 2. Statistical Distribution of Information Transmission Rate in Bits Per Second

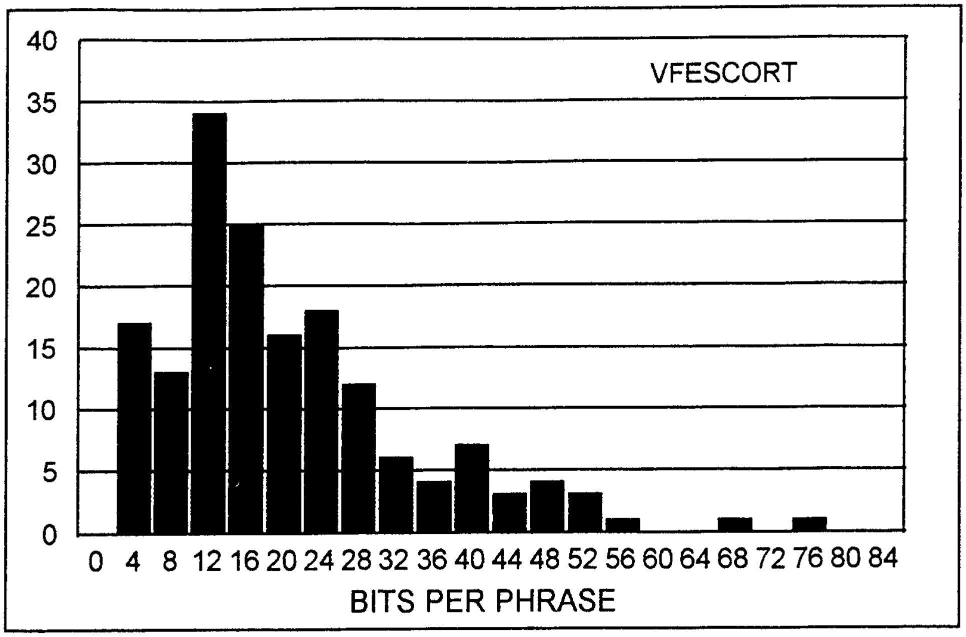


Figure 3. Statistical Distribution of the Amount of Information in Bits per Phrase

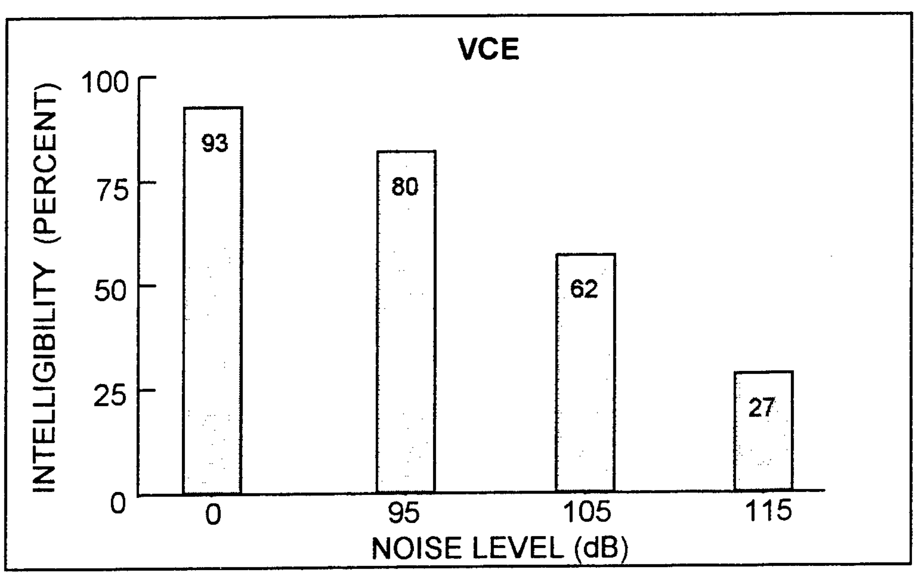


Figure 4. Representative Voice Communications Effectiveness Test Performance for A Range of Communications Conditions

JOINT ELECTRONIC WARFARE CENTER JAMMING RESEARCH RESULTS				
FILE NAME	NUMBER OF PRIMARY WORDS	NUMBER OF WORDS IN MESSAGE	ENTROPY OF PRIMARY WORD SET	AVERAGE MUTUAL INFORMATION OF MESSAGE
Sweep (JTIDS SWEEP)	238	908	6.8632	4.8361
LANEDEF (JTIDS-LANE DEFENSE)	233	1046	6.6499	4.6876
AWARE (JTIDS-THREAT AWARENESS)	204	661	6.6859	4.8698
TOTALS (JTIDS-TOTALS)	393	2615	7.0661	4.5646
VF ESCORT (NAVY VF ESCORT)	300	1192	6.9524	4.7192
TACEAGLE (NAVY TAC EAGLE)	285	1254	6.8914	4.5175
FLTTAC (NAVY FLT TAC)	206	1026	6.4824	4.7456
MARSHL (NAVY MARSHALL)	722	15923	6.5616	3.2289
VFCORD (NAVY VF CORD)	274	1222	6.7251	4.5707
APNET (NAVY APPROACH & NET)	627	14287	6.3797	3.7466
CAPSTA (NAVY CAPSTA 3)	205	844	6.5611	4.3100
STRIKE (NAVY STRIKE)	957	10645	7.3957	3.8353
LLD (NAVY LANDING LAUNCH DEPART)	1214	18221	7.1192	3.3632
NAVYTOLS (NAVY TOTALS)	2089	64614	7.3291	3.2671
WP 1086 (4 CHANNELS RECORDED)	608	5676	7.5155	4.4758
WP2	540	7656	7.1608	4.2317
WP3	415	3284	7.0855	4.4841
WP4	527	7248	7.2359	4.3637
WP5	437	5598	7.1196	4.4048
WP6	357	3962	6.8411	4.2747
WP7	375	4833	6.8556	4.4682
WP8	423	3566	7.1168	4.7359

Table 1. Mutual Information Analysis of Transcribed Speech Data That Formed the Basis of the Voice Communications Effectiveness Test

NUMBER OF WORDS	NUMBER OF PHRASES	SAMPLE PHRASES
2	30	NOT CLEAR
3	19	ALL ARE SOUTH
4	24	GOOD LUCK NEXT RUN
5	17	THREE STAY TEN MILES WEST
6	19	CLIMB BACK ONE DID NOT WORK

Table 2. Sample VCET Vocabulary Items Formed into Two to Six Word Phrases

1. marked	marsh	marks	mark
2. blast	fast	past	last
3. cone	code	cove	cold
4. reached	reach	reef	reads
5. scan	can	span	plan
6. seemed	seals	seems	seized
7. we	free	be	see
8. mapped	match	map	matched
9. parts	park	parked	part
10. real	she'll	wheel	fee
11. juts	jump	judge	just
12. fire	prior	wire	tire
13. great	straight	gate	state
14. thank	bank	rank	yank
15. tight	tied	type	timed

Table 3. Partial List of Confusable Words Used with the Two to Six Word Phrase VCET Vocabulary Items

VOCAL AGITATION AS A PREDICTOR OF EMOTION AND STRESS.

M.H. Allerhand and R.D. Patterson
MRC Applied Psychology Unit.
15 Chaucer Road, Cambridge CB2 2EF, UK

1. SUMMARY.

This paper reports an application of a computational auditory model to measure vocal agitation in speech automatically, and to relate it to the preceived stress in recordings of pilots operating under adverse conditions. Results of a short–time correlational experiment show significant correlation ($r = 0.765$; $p < .001$) between measured and perceived vocal agitation. It is also shown that time-integrated vocal agitation corresponds well with perceived stress over a period of the order of 18s.

2. INTRODUCTION.

Several years ago at the request of D.R.A. Farnborough, the Applied Psychology Unit developed a computational model of auditory peripheral processing to use as a tool for analysing the auditory environments in aircraft. Broadly speaking, the purpose of the project was to develop a tool to augment the oscilloscope and the spectrum analyser for real time analysis of acoustic environments in aircraft – everything from engine noise and gear whine to auditory warnings and speech. The resulting auditory model focuses on a representation referred to as the "auditory image". Recently D.R.A. Farnborough funded a project at the Applied Psychology Unit to determine whether the auditory image model (AIM) could be used to monitor pilots' speech and assist in determining whether the pilots were under undue stress. Psychological stress is a phenomenon that develops over days, weeks, and months which is well beyond the time scale of the memory in AIM. From recordings, however, it was apparent that voice quality often changes when people are under stress. A component of the change in voice quality is what might be termed "vocal agitation" which does occur on a time scale that is measurable with an auditory model. In this paper we report results from a project where AIM was used to develop a model of vocal agitation which in turn was used to perform an experiment which shows that vocal agitation is correlated with perceived stress in the human voice.

The advantage of a vocal stress measure is that it is essentially non–invasive; it does not require the operative to be wired-up or otherwise distracted in any way. We have developed a prototype Vocal Agitation (VA) monitor and performed evaluation experiments to demonstrate that VA measures can be used to monitor the level of arousal and emotional stress in the speaker. VA is a measurable quantity cued mainly by short–time pitch jitter; fluctuations in the glottal period over a sequence of 8 to 16 periods.

Early work on speech and the emotional states of pilots (Williams and Stevens, 1969, 1972) produced data which suggested some useful acoustic correlates of emotional states, but the work at that time was unable to specify any quantitative, automatic procedure that would reliably reveal the emotional state of a pilot. We are now able to specify a procedure to indicate emotional stress based on an objective measure of vocal agitation (VA).

We hypothesise that repeated or prolonged periods of VA build a picture of a stressed person over a suitably long time scale. According to this hypothesis emotion and stress would be indicated by a time integration of VA. Below we describe an experiment designed to correlate the VA measure with human judgements of the amount of stress in speech stimuli.

The method of extracting VA from speech is based upon a representation called the spiral mapping of the auditory image. We use this representation because it presents pitch jitter information in an explicit form, and because it can be normalised so that the resulting measure is independent of loudness, it rejects ambient noise, and it is speaker independent. The actual VA measure is a function of this representation, which is found by training on a corpus of stressed speech (Williams and Stevens, 1981).

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

3. THE AUDITORY IMAGE MODEL.

The auditory image model (Patterson, Allerhand, Giguere, 1995) analyses sound into a multi–channel representation using an auditory filterbank followed by a process of strobed temporal integration (STI). The STI is similar to a running autocorrelation of the filterbank output (or correlogram), but differs in two respects. Firstly, the samples which contribute to the running average at each autocorrelation lag are selected using a peak–picking mechanism (ie. not *every* sample is used). Secondly, there is no pairwise correlation: the output representation is simply the average value of lagged samples (ie. not the average value of lagged products). Besides the obvious computational advantages, these differences affect the properties of the output representation, although it is broadly similar to a conventional autocorrelation. The effect of using the average of lagged samples tends to compress the range of the output representation, which is useful in an auditory model. The main effect of the peak picking, however, is that periodic asymmetric patterns in the time–course of the input are preserved in the output representation.

asymmetric sounds, (like sinusoids with damped and ramped exponential envelopes), are perceptually very different despite having identical power spectra. This leads him to argue that asymmetry must be preserved in the time–integrated representation. Experiments with damped and ramped tones and noises show that perceptual differences are not sufficiently explained by asymmetries introduced by cochlear processes (Irino and Patterson, 1996), so that there is probably asymmetry enhancement during temporal integration.

3.1 The Spiral Mapping of the Auditory Image.

The spiral mapping of the auditory image reorganizes the information by wrapping the rectangular auditory image into a spiral. Periodic parts of the signal due to successive pitch periods, which were distributed across the rectangular auditory image as peaks at lags corresponding to successive multiples of the period, are brought into proximity on the spiral map, along radial "spokes" of the spiral (see figure 1).

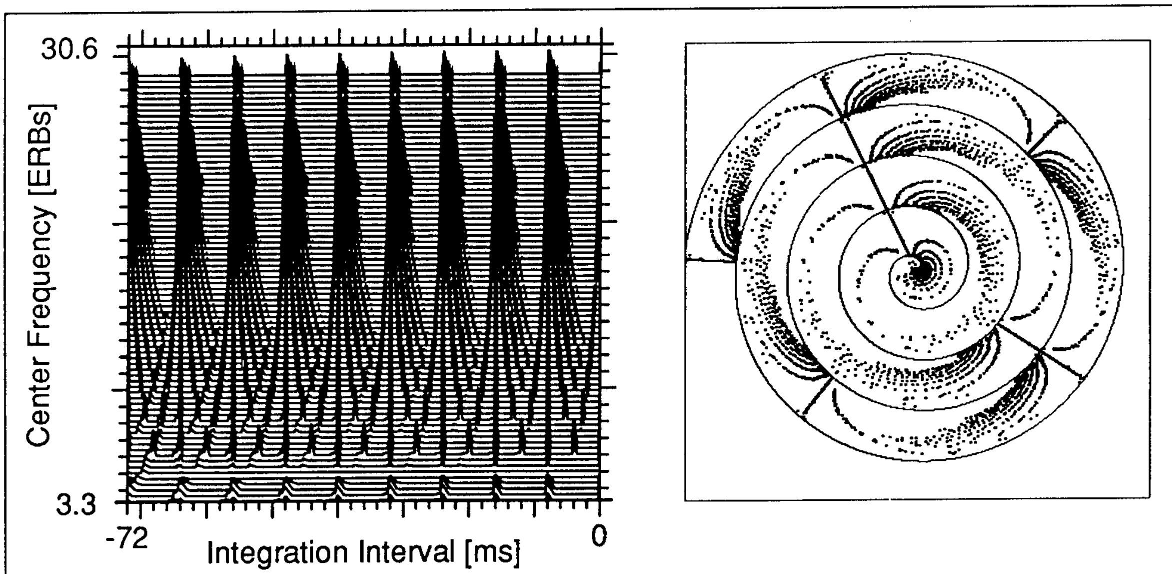


Figure 1. A rectangular and a spiral auditory image.

A conventional autocorrelation makes such patterns appear symmetric in the output. Patterson (1994) uses this property to justify STI as a model of temporal integration. He notes that temporally

The straightness of these spokes makes pitch jitter explicit and easier to detect. We can show that any periodic signal, regardless of the fundamental period, maps onto the *same* pattern of spokes

on the spiral map. This characteristic "pitch pattern" consists of the set of radial alignments of the peaks which correspond to successive multiples of the fundamental period; they appear as spokes on the spiral plane. The set of spokes have fixed and known angles in relation to each other, and this internal structure of the pitch pattern is independent of the fundamental period. This invariance property greatly facilitates pattern recognition based on the spiral map. The discrete spiral mapping $f(nT) \longrightarrow g(r_n, \theta_n)$ is a one-to-one mapping of the parameters, from the continuous time parameter (with sample interval T, producing the discrete signal f_n, $n = 0, 1, 2, \ldots$), to the polar coordinates (r_n, θ_n), representing a point in the spiral plane. The spiral is designed using a base–2 logarithm so that each cycle around the spiral corresponds to a doubling of the time period of the input. The mapping is defined:

$$\theta_n = 2\pi \log_2(nT) \tag{1}$$

$$r_n = \begin{cases} \log_2(nT) & \text{Archemedian form} \\ nT & \text{Logarithmic form} \end{cases} \tag{2}$$

where $n = 1, 2, 3, \ldots$. (Note: the sample index starts at 1, otherwise there would be an infinite number of spiral circuits within the first sample interval, since $\theta \rightarrow -\infty$ as $t \rightarrow 0$). Any doubling of the argument nT leads to an increase in θ_n of 2π radians (1), a complete cycle of the spiral. So points in time separated by a period which increasingly doubles have the same angular orientation on the spiral, and are aligned along a radial line or spoke of the spiral. In particular, if the input contains periodic peaks with period kT, synchronised with the origin so that the peaks occur at time nkT, $n = 1, 2, 3, \ldots$, then those peaks which occur at $kT, 2kT, 4kT, 8kT, \ldots$ (ie. with successively doubled period) form a spoke with angular orientation $2\pi \log_2(kT)$ radians. This first spoke of the pitch pattern is based on doublings of the fundamental period kT. However, the spiral map of a periodic signal produces a family of spokes, and this contains *all* the multiples of the fundamental period. Each spoke of this pattern consists of those peaks which occur at successive doublings of an odd–numbered multiple of the period kT. We show this by factoring the sample index into doublings of odd components using the following identity on the sequence of positive integers (proved in the appendix):

$$n \in \{1, 2, 3, 4, \ldots\} = \{2^p(2q+1) \mid p = 0, 1, 2, \ldots;\ q = 0, 1, 2, \ldots\}$$

So periodic peaks which occur at nkT can be factored into peaks which occur at $2^p(2q+1)kT$. Here the factor 2^p, $p = 0, 1, 2, \ldots$, generates the set of doublings of each odd numbered multiple of the period: $(2q+1)kT$, $q = 0, 1, 2, \ldots$. Substituting this into (1), and considering the map at periodic intervals of kT, we have:

$$\begin{aligned} \theta_{nk} &= 2\pi \log_2(2^p(2q+1)kT) \\ &= 2\pi p + 2\pi \log_2(2q+1) + 2\pi \log_2 kT \\ &= \theta_p + \theta_q + \theta_{kT} \end{aligned}$$

This shows the angle in terms of three components. Component $\theta_{kT} = 2\pi \log_2 kT$ is the orientation of the first spoke (based on interval kT), and is the primary angle for the pitch pattern as a whole. Component $\theta_q = 2\pi \log_2(2q+1)$ is the additional angle for each spoke, $q = 0, 1, 2, \ldots$, based upon an odd–numbered multiple of interval kT. Component $\theta_p = 2\pi p$ provides complete cycles of the spiral to align successive doublings of the intervals along radial spokes. Note that the primary angle for the pattern as a whole depends upon kT, but the internal structure of the pattern given by the additional angle for each spoke θ_q is independent of kT. So if the fundamental period kT varies, then the pattern as a whole rotates about the origin as the primary angle θ_{kT} varies, but the angles between the spokes of the pitch pattern remain constant. The form of the spiral (ie. Archemedian or logarithmic) affects the rate of increase of radius r_n; the angular structure of the pitch pattern is the same for both forms.

4. PATTERN RECOGNITION WITH THE SPIRAL AUDITORY IMAGE.

The spiral map is a representation of periodicity which highlights any small deviations such as might result from pitch jitter. These are seen as deviations in the straightness of the spokes of the pitch pattern (see figure 3.). The advantage of the spiral representation is that the pitch pattern has a fixed internal structure which is independent of the fundamental period. Provided the periodic peaks of the input are synchronised with the origin, then periodic input (with any period) forms a spiral map with straight spokes with fixed and known internal angles. This condition is met when the input is derived from an autocorrelation or STI process, since the origin is the peak at the zeroth lag. So pitch jitter is measured in terms of the straightness of the spokes of the pitch pattern.

We think of the pattern on the spiral map as a whole. We do not search the pattern for individual spokes, but we derive a transformation of the

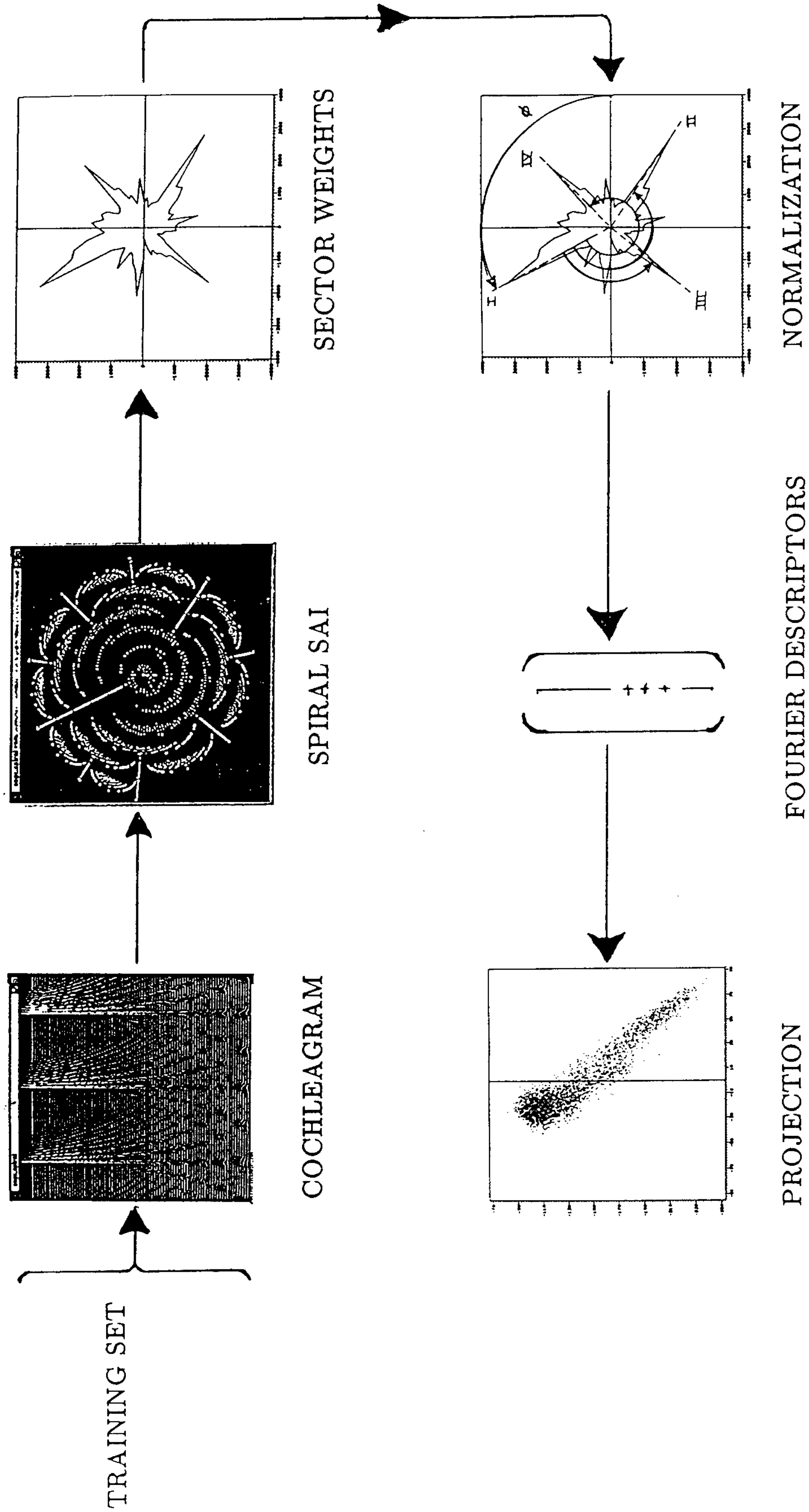
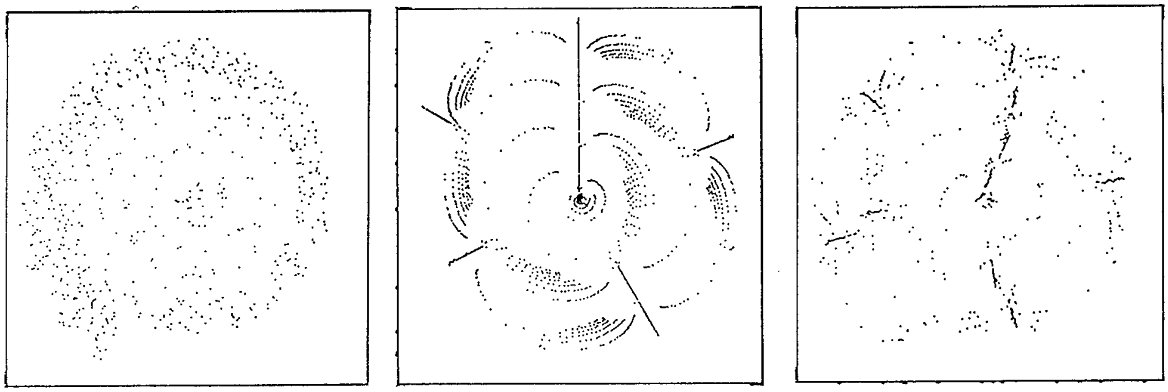


Figure 2. Training Procedure.

whole spiral onto a continuous scalar measure. We begin with a summary representation of the spiral map called the sector-weights contour (see figure 2). This is a one-dimensional summary of the two-dimensional map, constructed by dividing the spiral map into sectors. Each sector is represented by the sum of activity within it, and the sector-weights contour is the sequence of these values around the spiral map. This contour can be plotted as a closed polar contour, in which each value is plotted as a radius, to show the correspondence between large values and the sum activity produced by a spoke within a sector. Periodic information is now encoded in the shape of this contour. For example, a noise pattern which is a random pattern on the spiral map produces an irregular circular sector-weights contour. Contours of pitch patterns have a characteristic stellated structure, produced by spokes in fixed angular relationship to each other. We measure variation in the contour shape using the Fourier descriptors of the sector weights contour. These are the complex Fourier coefficients of the contour when we take the sector-weights map to be a complex plane. The use of Fourier Descriptors is a standard method of analysing silhouette shapes in pattern recognition (Wallace and Wintz, 1980). It enables a normalization of the scale and orientation of the sector-weights contour. Normalizing the scale of the contour makes the resulting measure independent of the size (ie. the sound level) of the input. Normalizing the orientation of the contour makes the resulting measure independent of the fundamental period of the input.



Noise. **Strong Pitch.** **Pitch Jitter.**

Figure 3. Spiral auditory image showing the effect of pitch jitter.

4.1 Vocal Agitation Measurement.

One of the correlates of VA is pitch jitter. Whereas unagitated voiced speech is characterised by reasonably solid spokes on the spiral map, agitated voiced speech shows jittery spokes, caused by small local pitch variations (see figure 3). The VA measure is a continuous measure of the difference between these two types of pitch pattern.

The training procedure is to find the projection from the pattern space which discriminates a set of training stimuli. The training stimuli were produced by actors who could work themselves into a convincing emotional state (Williams and Stevens, 1981). A training set for two pattern classes (agitated and unagitated) was constructed by selecting frames from this recorded speech which were labelled respectively as "angry" and "normal". These training stimuli were processed by the auditory image model so as to create a distribution of training vectors in the pattern space of the spiral auditory image (as described above). The VA measure was defined as the normal to a linear discriminant for the two pattern classes. The linear discriminant was trained using the method of optimal discriminant planes (Foley and Sammon, 1975). The VA measure is the length of the projection onto the normal to this linear discriminant. This length measures the distance of any projection along a line between the class "unagitated" and the class "agitated".

5. TIME–INTEGRATED VA AS A STRESS MEASURE.

Williams and Stevens (1981) found that speech produced under conditions of emotional stress contains some words or phrases which, when heard in isolation, tend to be judged to originate from a stressed speaker, but some others which do not. Words or phrases which sound as if produced by someone under stress are examples of periods of VA. The VA during these periods is easily detected by humans, and is measurable. However, there are also periods during a speech produced by a person under stress which do not sound vocally agitated. This observation has lead us to the hypothesis that repeated or prolonged periods of VA build a picture of a stressed person over a suitably long time scale. According to this hypothesis emotion and stress would be indicated by a time integration of VA. Given an extended passage of stressed speech we would predict that human judgements of the stress in isolated words and phrases should vary over a range of levels, and that these should correlate with the VA measured for the words and phrases. Over a longer time scale (eg. several phrases) we would predict that human judgements of the stress level should converge onto a single integrated decision, and that this should correlate with the VA measure integrated over this time period. This hypothesis is tested in the following experiment.

5.1 Method.

An experiment was designed to correlate the VA measure with human performance in a perceptual judgement task to rate the amount of stress in given speech stimuli. The source of the speech was a continuous cockpit recording lasting for some 10 minutes of two pilots flying a Hunter aircraft over England and Wales. The speech in the recording was mixed with mask noises such as the high–level breathing noise produced when the pilots experience "G" forces. The early part of the flight is routine and the pilots seem relaxed. During the later part of the flight the pilots become excited (and stressed) when other aircraft are unexpectedly encountered and they engage in mock combat.

The stimuli were 63 phrases, each between 2 and 5 seconds duration, which were excised from the continuous cockpit recording. The phrases were presented to five subjects in random order. The subjects were asked to rate the amount of stress they perceived in each phrase on a 5-point scale. They were instructed to base their judgements on the sound of the voices and to try not to be influenced by the meaning of the words. Each subject was presented with three training phrases representative of minimal stress (score=0), medium stress (score=3), and high stress (score=5), and they were encouraged to re-train themselves at intervals during the experiment. VA measurment was made for each of the 63 phrases. The raw VA measure fluctuates rapidly during a phrase, varying with speaker rate and the rates of word onsets and offsets. The maximum VA measure during the phrase was taken as the VA value of the phrase.

In a second experiment, the subjects were presented with the continuous 10 minute recording and asked to rate the vocal stress. The pattern of the subject's account was compared with the smoothed VA measure over a range of time–constants.

5.2 Results.

The results of the first short–time correlational experiment are shown in figure 4, a scatter plot of the pooled perceptual data against the VA measure. Both axes have been normalized for zero mean and unit standard deviation in the respective distributions. A correlation analysis of the pooled experimental data with the maximum VA value for each of the 63 phrases showed a significant correlation ($r = 0.765$; $p < .001$).

For the second experiment, the subjects all reported that the early part of the flight seemed fairly relaxed, but that the level of stress increased markedly during the later part of the flight, centering around two periods of maximum stress. The time-integrated VA measure can be seen in figure 5. It shows the results of smoothing the short-time VA measure by convolution with a Gaussian mask, over a range of mask variances. It can be seen that as the integration time (controlled by the variance of the mask) is increased, the rapid fluctuations in the VA measure are smoothed out, tending towards a pattern which shows relatively little activity at the start, and two bursts of activity during the latter half. This pattern corresponds with the subject's account of the stress, and also with two encounters with other aircraft which occurred during the flight. Figure 5 suggests that a Gaussian smoothing mask with at least 10s duration is required to reveal the stress level underlying VA. This corresponds to an integration period of the order of 18s (ie. 6 standard deviations).

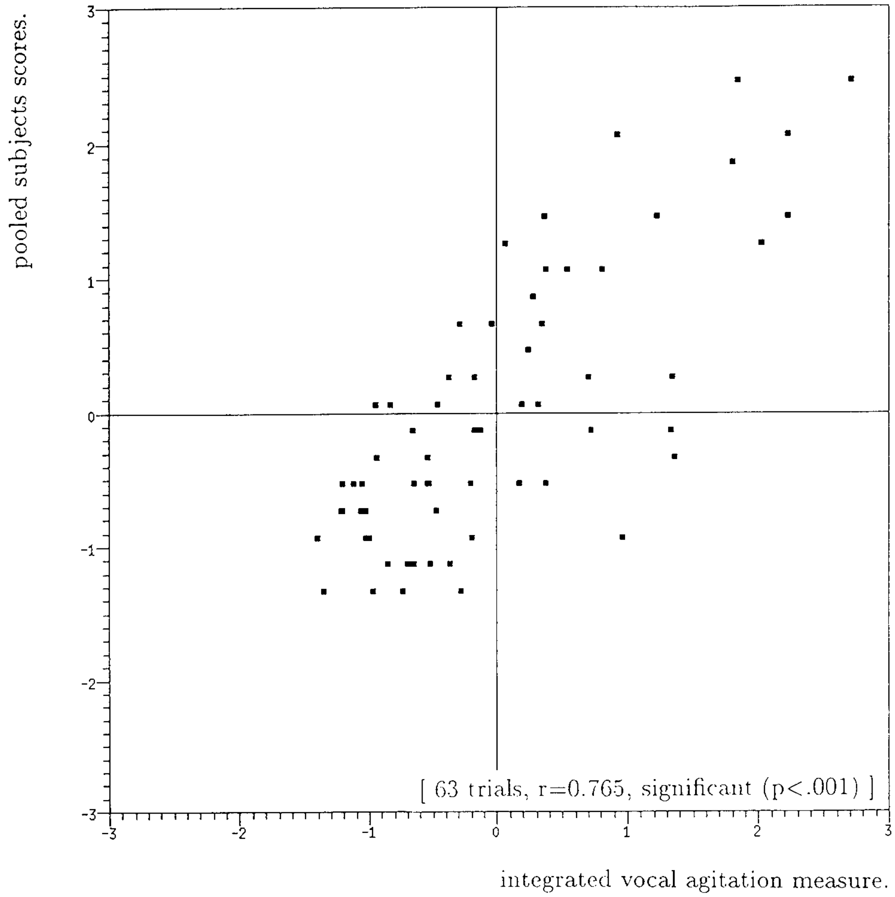


Figure 4. Scatter plot of perceived vocal agitation against VA measure.

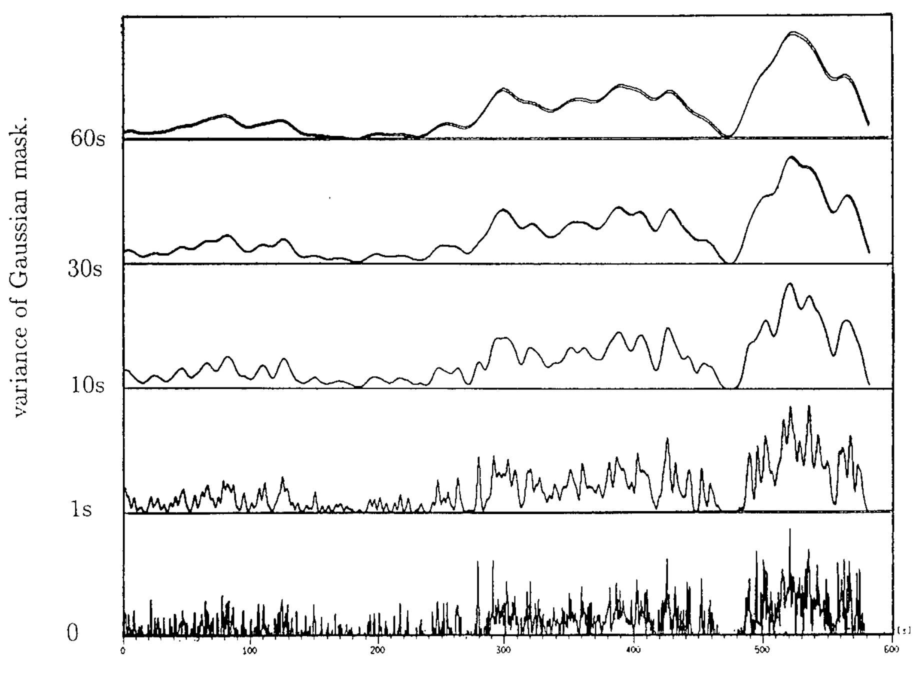


Figure 5. Vocal agitation over a range of scales on integration.

5.3 Discussion.

It is important to distinguish between VA and stress. Emotion and stress are subjective psychological states which give rise to involuntary physiological responses which produce VA in speech. Sporadic periods of VA are a symptom of stress. Unlike the galvanic skin response, for example, vocal stress can be consciously suppressed. However unless a deliberate effort is made, the vocal effects of stress can be heard even from those who are trained to keep calm under difficult conditions. It is well known that stress can cause a physiological response. We suggest that this leads, amongst other things, to a tightening of the muscles surrounding the vocal folds, and that this changes the mechanical properties of the glottis. One of the results of this is that voiced (ie. quasi-periodic) speech sounds have a certain quality which people judge as sounding "agitated", which seems mainly to be caused by pitch jitter. While VA fluctuates quite rapidly, for example within words or from word to word, it is assumed that the underlying psychological state is relatively slowly varying. For example, in a perception experiment on identifying vocal stress (Williams and Stevens, 1981) method actors were asked to produce a passage of "angry" speech, which they did by working themselves into a sustained stressed condition. Sentences were excised from the passage of angry speech and presented in isolation to subjects for judgement as to the level of vocal stress. It was found that the stressed speech contained some words or phrases which sounded angry, but some others which did not. Words or phrases which sound angry are examples of periods of VA. The VA during these periods is easily detected by humans, and is measurable. However, there are also periods during a speech produced by a person under stress which do not sound vocally agitated. This observation has lead us to the hypothesis that repeated or prolonged periods of VA build a picture of a stressed person over a suitably long time scale. According to this hypothesis emotion and stress would be indicated by a time integration of VA. Given an extended passage of stressed speech we would predict that human judgements of the stress in isolated words and phrases should vary over a range of levels, and that these should correlate with the VA measured for the words and phrases. The results of the short-time correlational experiment showed that people's judgement of the stress level in words and short phrases varied widely as predicted and in agreement with other similar experiments (Williams and Stevens, 1981). Over a longer time scale (eg. several phrases) we would predict that human judgements of the stress level should converge onto a single integrated decision, and that this should correlate with the VA measure integrated over this time period. The subjects demonstrated an ability to integrate the fluctuating short-time vocal agitation into a more stable perception of stress over a longer time period. This supports the assumption that the underlying psychological state is relatively slowly varying.

The second experiment suggests that integrated VA corresponds with the stress which the subjects perceived in the speech passage over a longer time scale. The integration period required to smooth the VA measure in order to correspond with the subject's account suggests that the underlying psychological state which gives rise to periods of VA (interspersed with periods of non-VA) will produce sufficient VA over a period of about 18s to enable both humans and measurements to determine the level of stress.

5.4 Conclusions.

The hypothesis that the perception of psychological stress in speech is cued by a time-integration of short-time vocal agitation is supported by a comparison between the short and long-time experiments. This may explain why experiments to determine the degree of stress in short stretches of speech excised from a stressed passage were inconclusive (Williams and Stevens, 1981). The correlations obtained between perceived stress and VA (in the short-time case) and integrated VA (in the long-time case) show that the VA measure is very accurately predicting the level of vocal agitation even in the presence of mask noise in the cockpit recording, and that a time-integration of the VA measure is an indicator of subjective psychological stress. The integrated VA measure accurately predicted the stress which the subjects perceived in the speech passage over a longer time scale of the order of about 18s.

6. APPENDIX.

To prove the following identity on the set of positive integers:

$$\{1,2,3,4,\ldots\} = \{2^p(2q+1) \mid p=0,1,2,\ldots;\ q=0,1,2,\ldots\}$$

1. The factor $(2q+1)$, $q=0,1,2,\ldots$ represents the set of all odd numbers, $\{1,3,5,7,\ldots\}$.

2. Every even number is a doubling of either an odd number or an even number. It follows that every even number is a member of a set $\{2^p a \mid p = 0, 1, 2, \ldots\}$, where a is an odd number, and the factor 2^p, $p = 0, 1, 2, \ldots$ represents a succession of doublings. The union set of all doublings of all the odd numbers, $\{2^p(2q+1) \mid p = 0, 1, 2, \ldots;\ q = 0, 1, 2, \ldots\}$, must then contain all the positive integers at least once.

3. To show that the succession of doublings of each odd number generates a *unique* set of even numbers, it is sufficient to show that the sets $\{2^{\pm p} a \mid p = 0, 1, 2, \ldots\}$, and $\{2^{\pm q} b \mid q = 0, 1, 2, \ldots\}$, are disjoint for all $a \neq b$, where a and b are odd numbers. If $2^p a = 2^q b$, then $a = 2^{q-p} b$. But this can only be odd in the trivial case of $p = q$, when $a = b$. Similarly, if $2^{-p} a = 2^{-q} b$, then $a = 2^{p-q} b$, and again this can only be odd in the trivial case of $p = q$, when $a = b$.

4. It follows from the above that the set $\{2^p(2q+1) \mid p = 0, 1, 2, \ldots;\ q = 0, 1, 2, \ldots\}$ contains all the positive integers once only, and is identical to the set of positive integers $\{1, 2, 3, 4, \ldots\}$.

7. REFERENCES.

Allerhand M., Patterson R., Robinson K., Rice P. (1991) "Application of the SVOS Algorithm", *Spiral VOS Interim Report V*, MOD PE SLS/42B/663, August 1991.

Foley D., Sammon J. (1975) "An Optimal Set of Discriminant Vectors", *IEEE Trans Comput*, **24**, 281–289.

Irino T., Patterson R., (1996) "Temporal asymmetry in the auditory system," J. Acoust. Soc. Am. 99, 2316-2331.

Patterson R. (1994) "The sound of a sinusoid: time–interval models" *J. Acoust. Soc. Am.* 96 1419–1428.

Patterson R., Allerhand M., Giguere, C., (1995). "Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. 98, 1890-1894.

Wallace T., Wintz P. (1980) "An Efficient Three-Dimensional Aircraft Recognition Algorithm using Normalized Fourier Descriptors", *Computer Graphics and Image Processing*, **13**, 99–126.

Williams C.E., Stevens K.N. (1969) "On determining the emotional state of pilots during flight: An exploratory study". *Aerospace Medicine* 40 1369–1372.

Williams C.E., Stevens K.N. (1972) "Emotionals and speech: Some acoustical correlates". *J. Acoust. Soc. Am.* 52 1238–1250.

Williams C.E., Stevens K.N. (1981) "Vocal Correlates of Emotional States". In: Speech Evaluation in Psychiatry, J.K. Darby (ed), Grune and Stratton, Inc..

EFFECTS OF HELICOPTER CABIN NOISE UPON HF VOCODER SECURE SPEECH SYSTEMS

I.E.C. Rogers
Dr G.M. Rood
Systems Integration Department, Air Systems Sector
Defence Research Agency, Farnborough
Hampshire GU14 6TD, UK

SUMMARY

An increasing number of military aircraft are being provided with secure (encrypted) systems for air-to-air, air-to-ground and ground-to-air communications.

Most secure HF radio channels use a Linear Predictive Coder (LPC-10 Vocoder) to parameterise the talker's speech, and this digital data is then encrypted before being transmitted over the HF radio link. At the receiver, the data is decrypted and fed into a second vocoder, where the speech parameters transmitted are used to produce a representation of the original speech signal. The vocoders transmit the digitised data at 2.4kbits/sec according to the NATO STANAG 4198 interoperability standard [1].

Studies at DRA Farnborough have identified that the presence of helicopter noise at the microphone input to the transmitting vocoder reduces the intelligibility of the vocoded speech transmitted, and that the reduction is dependent on the relative levels of the speech and noise at the microphone (i.e. the speech to noise ratio, SNR). These assessments have been conducted using Diagnostic Rhyme Test (DRT) techniques.

DRA have investigated techniques to enhance the performance of vocoders using digital processing techniques. DRT and user acceptability assessment trials have been conducted to assess the effects of these techniques on LPC-10 vocoder performance and the results of this work will be presented.

1. BACKGROUND

1.1 LPC-10 vocoders

Figure 1 depicts the model of speech production employed by LPC vocoders [2]. In this model, random noise excitation is selected when the speech is unvoiced and periodic excitation is selected when the speech is voiced. In addition, the pitch period is estimated when the speech is voiced. The pitch period varies between 20ms for a deep male voice and 2ms for a high pitched child or female.

The gain stage of the model in Figure 1 controls the amplitude of the output, in order to provide a close approximation to real speech. The output of the gain stage is passed through a vocal tract filter, which models the frequency characteristics of the vocal tract. LPC vocoders model the vocal tract as an all-pole digital filter. In particular, LPC-10 vocoders derive ten

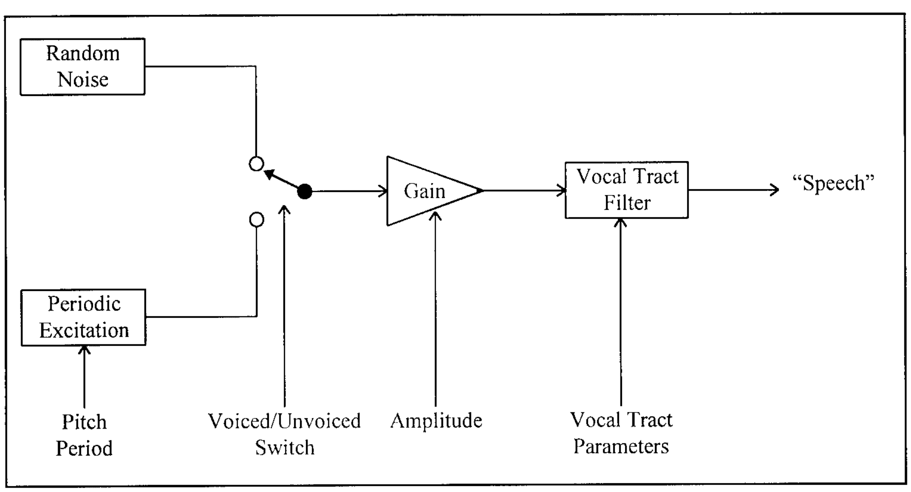


Figure 1 - Speech Production Model

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

filter coefficients and therefore five vocal tract poles are modelled. The vocal tract parameters are modified by the movement of the speech articulators (tongue, jaw, lips, etc), and thus must be updated, typically at intervals of between 5ms and 25ms.

In summary, the speech signal is parameterised into:

- voiced/unvoiced excitation decision;
- pitch period (if voiced);
- gain;
- vocal tract filter coefficients.

The original speech signal can be synthesised at a second vocoder if these speech parameters are transmitted to it.

1.2 Secure HF radio channels

Figure 2 depicts a typical secure HF channel. The speech signal from the operator's microphone is routed to an LPC vocoder where it is parameterised. The parameters are transmitted as a digital bit stream to the Encryptor and the encrypted data is then processed by an HF modem and transmitted.

At the receiving radio, the modem converts the signal back to a digital bit stream and routes it to the decryptor. A second vocoder synthesises the original speech signal from the decrypted data.

To provide inter-operability between NATO platforms vocoders compliant with NATO STANAG 4198 [1] are required. This STANAG specifies the use of LPC-10 vocoders with an analysis frame length of 22.4ms and with 54bits/frame (i.e. a data rate of 2.4kbits/sec).

Note that higher bit rate vocoders are available for VHF and UHF communications, but are unsuitable for HF channels since these are limited to a 3kHz bandwidth.

1.3 The helicopter noise environment

Narrowband and 1/3 octave band noise spectra for the helicopter cabin noise used for the studies described in this paper are shown in Figures 3 and 4. The overall noise level is 104.8dB SPL (90.5dB(A) SPL).

The noise is a combination of tonal and broadband components and is predominantly low frequency. The major noise sources are mechanically-induced noise from the gearbox, transmission train and rotors and aerodynamically-induced noise.

1.4 Talker speech-to-noise ratio

Vocoder performance is dependent on the speech-to-noise ratio (SNR) of the talker. SNR is the ratio of the speech level to the background noise picked up by the talker's microphone. DRA have conducted ground simulations and flight trials to measure the SNR values achieved by a range of aircrew under various helicopter noise conditions. Table 1 tabulates the ranges of SNR values achieved in each these trials.

Trial Type	SNR Range
Flight Trial	5 - 28dB
Ground Simulation	12 - 35dB

Table 1 - Talker SNR in Helicopters

Table 1 shows that large ranges of SNR were measured during the two trials. The main cause of this was inter-subject variability in vocal effort. It was noted, however, that deviations in background noise level had a smaller effect on SNR.

2. PERFORMANCE ASSESSMENT METHODS

2.1 Diagnostic Rhyme Test

The Diagnostic Rhyme Test (DRT) is widely used to assess the intelligibility of voice communications systems and has become a NATO standard for assessing linear predictive coders [1]. A detailed account of the development of the DRT is given by Voiers [3]. The DRT is based on the ability of a listener to distinguish between pairs of words which differ only in one acoustic attribute of their initial consonant. There are 192 words arranged in 96 rhyming pairs in the DRT vocabulary. For example, "veal" and "feel" are a rhyming pair which differ because the initial consonant is voiced in "veal", but unvoiced in "feel". The six attributes tested are voicing, nasality, sustention, sibilation, graveness and compactness. The complete DRT vocabulary is reproduced in Annex A.

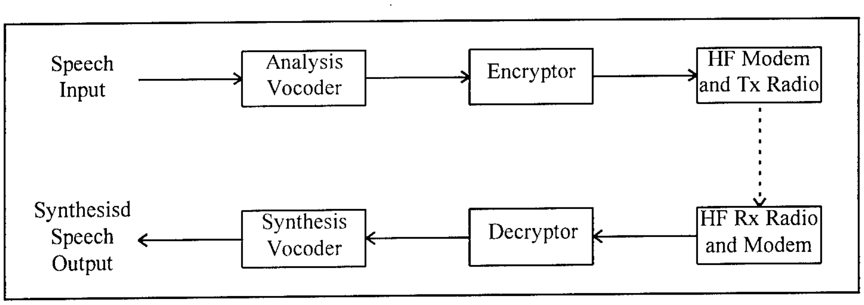


Figure 2 - Typical Secure HF Radio Channel

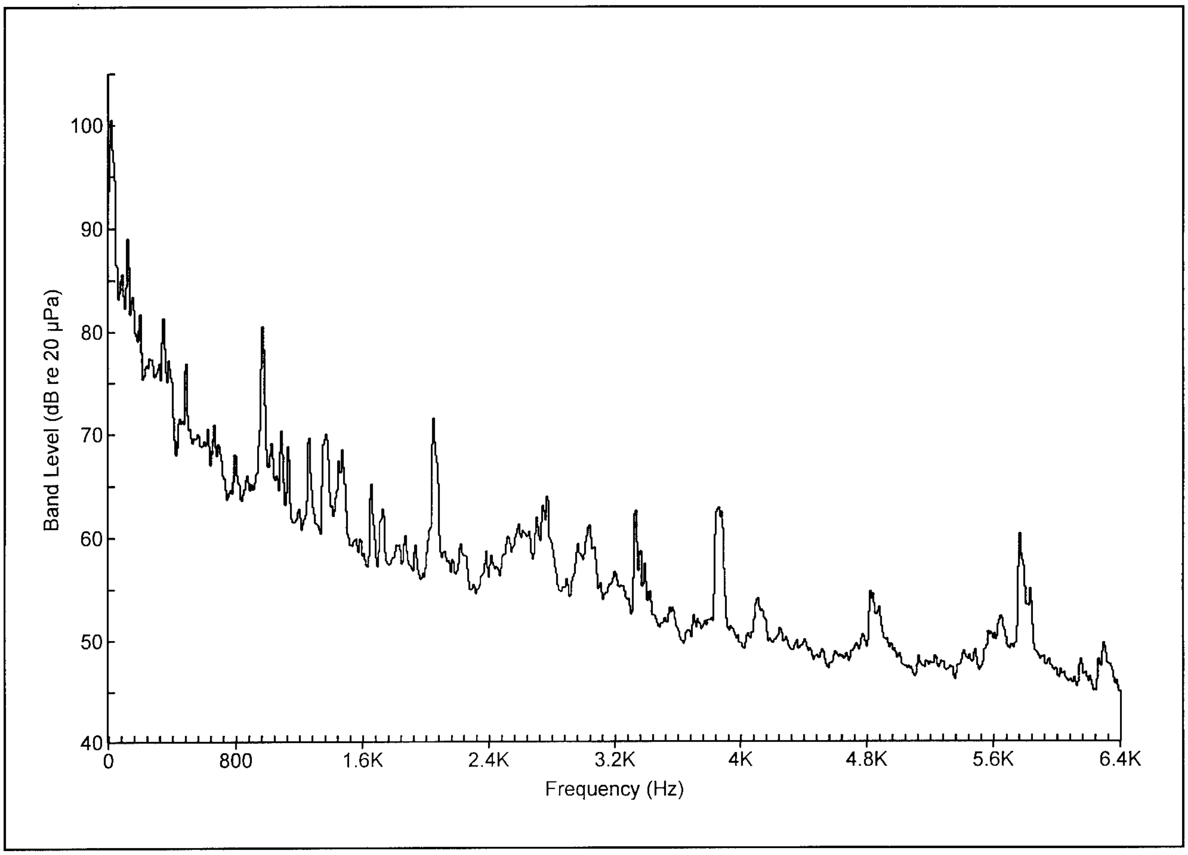


Figure 3 - Narrowband Helicopter Noise Spectrum

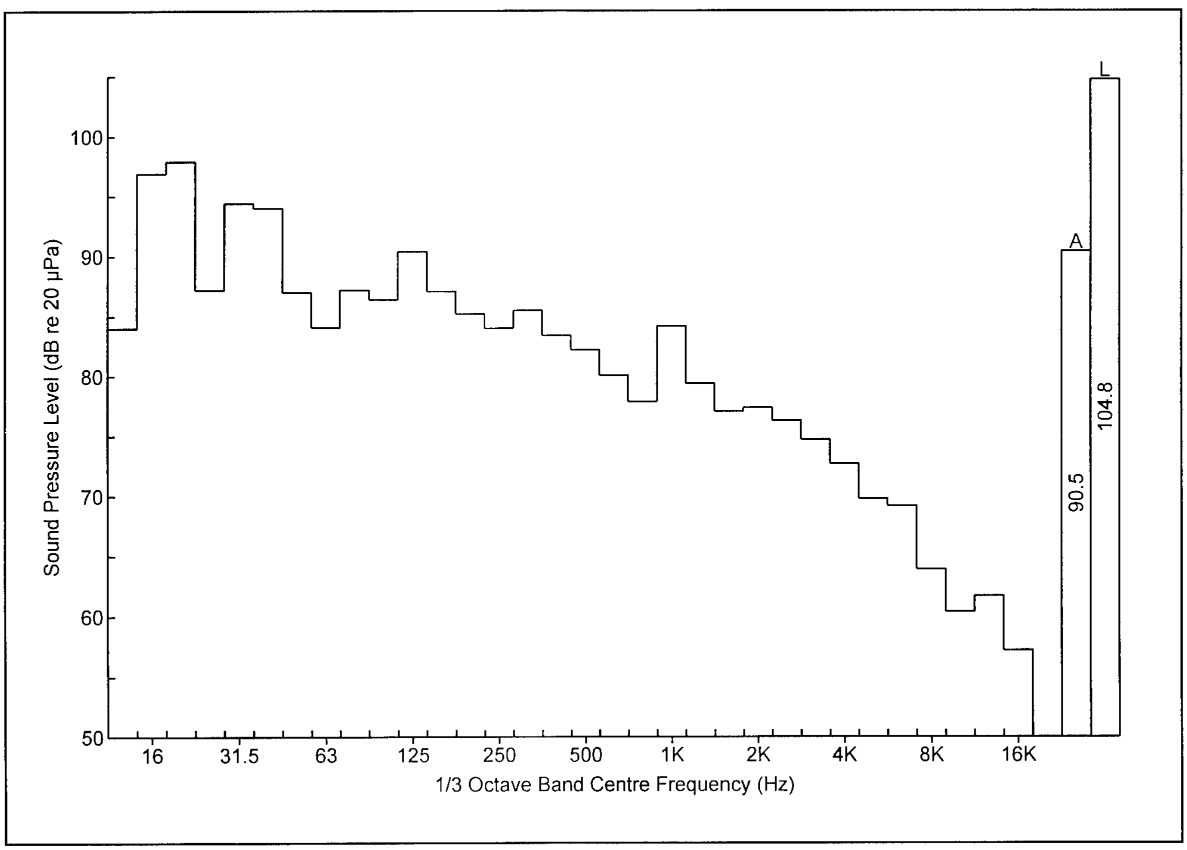


Figure 4 - 1/3 Octave Band Helicopter Noise Spectrum

DRT tests always use pairs of lists called "half-lists". The first list contains one word from each of the 96 word pairs, and the second list contains the other 96 words. This means that the tests are "balanced" because the pair of half-lists contains each of the 192 words once. There is an enormous choice of which word from each word pair appears in each half list, but in practice 26 standard half-list pairs are used.

The result of a DRT test is expressed as a percentage of correct responses, corrected for guessing. This means that a listener who recognises half of the words correctly will score 0% as this result could have been achieved by guessing.

The validity of DRT results is highly dependent on the listening panel used. The panel must be audiologically screened to check their hearing. The listening panel must also be "trained" by completing a series of DRT training runs before the full tests are conducted. These are designed to identify the listeners who are not suited to the long periods of concentration necessary for the tests. They also allow checks to be made on the consistency and repeatability of the performance of individual listeners - an essential feature of the DRT.

Other techniques such as the Articulation Index (AI) [4] and Speech Transmission Index (STI) [5] are available to evaluate communications systems. However, they are not suitable for testing linear predictive vocoders [6].

2.2 User assessment

DRT scores provide a well-proven and repeatable method of measuring the intelligibility of communications systems. However, it is important to relate DRT scores to the results of user assessments of the communications systems. For example, a system might yield a high DRT score but might prove unacceptable to the user group because of the degree of effort actually required to communicate successfully using the system.

The user assessment tests conducted at DRA Farnborough are based on pre-recorded sentences, which are processed through the communications systems to be evaluated. In order to make the tests as realistic as possible, "contextual" sentences are used and these are based on messages that would be transmitted over the communications systems in-service.

The sentences are replayed to a panel of users, who are asked to rate them on the metrics shown in Table 2. These metrics are based on previous work reported at [7]. Note that subjects are allowed to select any point on the rating scale for the first three metrics, but can only select either "Acceptable" or "Unacceptable" for the Acceptability rating.

Criteria	Scale Range	Scale End Point Descriptors	
Intelligibility	1 - 10	Totally Unintelligible	Completely Intelligible
Quality	1 - 10	Extremely Degraded	Completely Natural
Effort Required	1 - 10	Extreme Effort	No Special Effort
Acceptability	0 or 1	Acceptable	Not Acceptable

Table 2 - User Assessment Metrics

2.3 Speech Intelligibility Facility

Tests are conducted in the DRA Farnborough Speech Intelligibility Facility [8]. This consists of a reverberant noise chamber equipped with a noise generation system to produce realistic background noise fields for the tests. The tests are administered from an adjoining control room.

The chamber is fitted with thirteen listener stations, each equipped with a Personal Computers (PC), response box (for DRT tests) and mouse (for user assessment tests). A master PC in the control room controls the conduct of the test and logs the subject responses for analysis.

The pre-recorded DRT word lists (for DRT tests) and contextual sentences (for user assessment tests) are presented aurally to the listeners using an audio replay system. The output from a Digital Audio Tape (DAT) recorder is played into a master replay unit, from where it is distributed to remote audio boxes located at the thirteen listener stations. The replay system is designed so that listeners may use appropriate aircrew helmets, headsets or headphones, depending on the communications system to be tested.

3. PREVIOUS STUDIES AT DRA

3.1 Vocoder comparison

A series of DRT tests were conducted on two NATO STANAG 4198 compliant LPC-10 vocoders in helicopter noise at various talker SNRs.

Both systems produced low DRT scores and hence poor intelligibility. However, it was found that one system performed significantly better than the other and therefore this has been used for all subsequent work at DRA.

3.2 Effects of noise on parameter determination

A computer interface and software was developed so that the digital data transmitted from the analysis vocoder to the encryptor (Figure 2) could be captured on a computer. This data stream was then analysed to derive the speech parameters transmitted by the analysis vocoder.

Speech at various SNRs was presented at the input to the vocoder and the digital data output by the vocoder stored on the computer. The transmitted speech parameters were studied to determine which parameters were most effected by background noise at the speech input.

The work demonstrated that the vocal tract filter coefficients and the voiced/unvoiced excitation decision were most affected by helicopter noise.

4. PRE-PROCESSOR STUDIES

The results described in Section 3.1 indicated that the intelligibility of STANAG 4198 vocoders was poor in helicopter noise. Therefore, UK MOD has partially funded two companies to develop and demonstrate prototype speech pre-processors.

The pre-processors are introduced at the speech input to the analysis vocoder (Figure 2). They are designed to analyse the communications microphone signal (speech + noise) to remove the noise component of the signal whilst leaving the speech component unaffected. The "cleaned up" speech signal is then input to the vocoder.

A selection process will be conducted by UK MOD to select one pre-processor for a production system. As part of this process, the prototypes have been evaluated at DRA Farnborough using DRT and user assessment techniques.

The detailed results of these tests are commercially-sensitive and therefore only the outline results will be presented here. Detailed results will be published in a future DRA report.

It is hoped that further pre-processor development under the production contract will mean that the production system will offer improved performance over the prototypes.

5. ASSESSMENT OF PRE-PROCESSORS

The pre-processors have been assessed using DRT and user assessment techniques, as described in the following sections. Note that for these tests, the equipment configuration shown in Figure 2 was simplified by connecting the encryptor and decryptor "back-to-back" and therefore the tests did not include the effect of the modems, radios or HF transmission path on communications performance.

5.1 DRT Tests and Results

DRT tests were conducted to identify the effect of each pre-processor on vocoder performance. The following noise conditions were tested:

a) 0, 5, 10, 15, 20 and 25dB talker SNR in helicopter noise, listener in helicopter noise.

b) Talker in quiet, listener in helicopter noise.

c) Talker and listener both in quiet.

Five talkers and eleven listeners participated in the tests.

Mk4 aircrew flying helmets and Racal 8956 microphones were used for all the tests, as these are commonly used by helicopter aircrew.

Table 3 presents the mean DRT results for the vocoder alone (i.e. without any pre-processor) and the results are plotted in Figure 5. The 0, 5, 10, 15, 20 and 25dB axis labels identify tests conducted at that talker SNR and with the listeners in helicopter noise. As expected, the DRT score increases as the SNR improves (increases). Note that at 0dB SNR the intelligibility was too poor to conduct a DRT test.

Talker Noise	Listener Noise	DRT Score (%)
0dB SNR	Helicopter	n/a
5dB SNR	Helicopter	36.5
10dB SNR	Helicopter	53.4
15dB SNR	Helicopter	59.2
20dB SNR	Helicopter	66.0
25dB SNR	Helicopter	66.7
Quiet	Helicopter	72.5
Quiet	Quiet	74.6

Table 3 - Mean DRT Scores for Vocoder

In Figure 5, the "Quiet" axis label identifies the test conducted with the talker in quiet and the listener in helicopter noise. The 72.5% DRT score at this condition represents the maximum performance that could be achieved by a pre-processor + vocoder system if the pre-processor worked perfectly (i.e. removed all the noise without affecting the speech signal). This would also approximate to ground-to-air communications (if the ground environment was quiet).

The "Quiet-Quiet" axis label identifies the test conducted with the talkers and listeners both in the quiet. The 74.6% DRT score at this condition represents the performance of a pre-processor + vocoder if the pre-processor worked perfectly, and there was no noise at the listener position. This would also approximate to ground-to-ground communications (in a quiet ground environment).

The results for the pre-processors are not included in Figure 5 for the reasons discussed in Section 4. However, in general the pre-processors increased the DRT scores by around 15% at SNRs below 15dB. At 0dB SNR, DRT scores in excess of 40% were achieved.

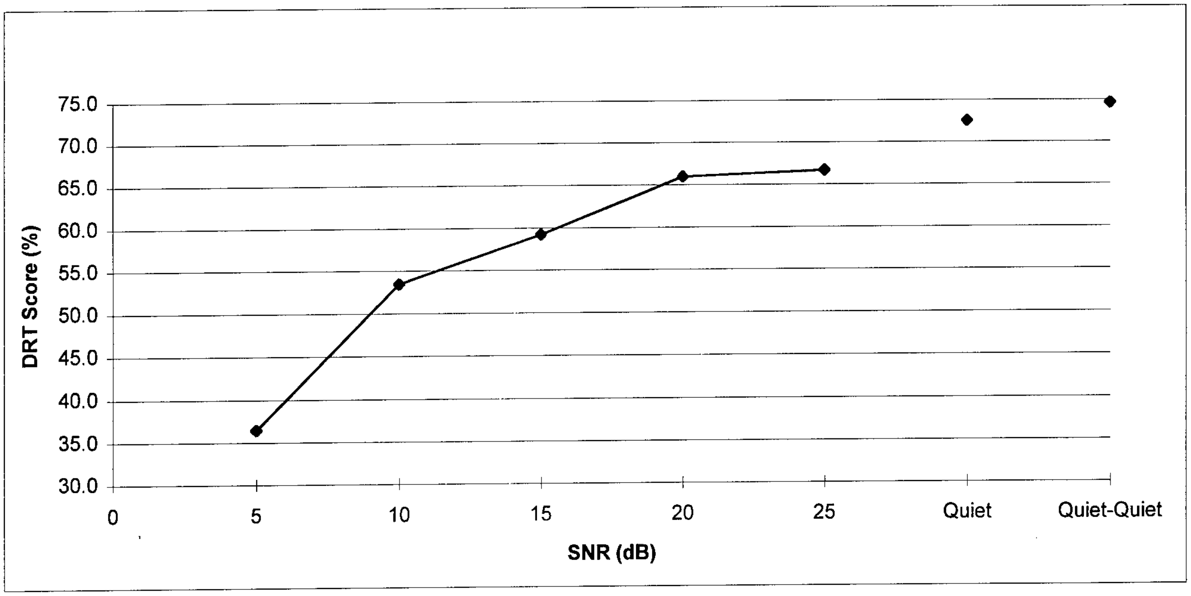


Figure 5 - DRT Results for Vocoder Only

5.2 User Assessment Tests and Results

User assessment tests were conducted to identify the effect of each pre-processor on vocoder performance. In addition, tests were conducted for "clear" speech (i.e. not vocoded) as a baseline condition. The following noise conditions were tested:

a) 0, 5, 10, 15, 20 and 25dB talker SNR in helicopter noise, listener in helicopter noise.

b) Talker in quiet, listener in helicopter noise.

c) Talker and listener both in quiet.

The user panel consisted of twelve experienced helicopter aircrew, who rated each passage using the metrics in Table 2.

Mk4 aircrew flying helmets and Racal 8956 microphones were used for all the tests, as these are commonly used by helicopter aircrew.

The ratings made by each of the 12 subjects were averaged to provide an overall rating for each system under each noise condition.

5.2.1 User assessment of intelligibility

The user assessment of intelligibility, averaged over the twelve listeners, is presented in Table 4 and Figure 6. The axis labels have the same meanings as those in Figure 5.

As expected, the users rated the "clear" speech channel as more intelligible than the vocoded speech at all noise conditions. The vocoded speech channel intelligibility rating increases with improving SNR, but the clear channel exhibits a flatter profile, except at the very low SNRs where the intelligibility rating decreases rapidly.

It is interesting to observe that even at the "Quiet" and "Quiet-Quiet" conditions the vocoded channel achieves an intelligibility rating of less than 7 points.

The results for the pre-processors are not included in Figure 6 for the reasons discussed in Section 4. However, below 15dB SNR the pre-processors increased user ratings of intelligibility by up to 1.9 points.

5.2.2 User assessment of quality

Table 5 and Figure 7 present the results for the user assessment of quality, which show the same general trends as the intelligibility assessment (Figure 6). Above 20dB there is no subjective improvement in the quality of the vocoded channel and the rating does not exceed 6 points at any condition.

The pre-processors increased the quality rating of the vocoded channel by up to 2.1 points at talker SNRs below 15dB.

5.2.3 User assessment of effort required to listen

Table 6 and Figure 8 present the results for the user assessment of effort required, which show the same general trends as Figure 6. Above 20dB there is no subjective improvement in the rating of the effort required to listen to the vocoded channel.

The pre-processors improved the rating of effort required to listen by up to 1.9 points at talker SNRs below 15dB.

5.2.4 User assessment of acceptability

Table 7 and Figure 9 show the percentage of users that rated the clear and vocoded systems as acceptable under each noise condition. The clear channel is rated as acceptable by all aircrew, except below 5dB talker SNR. However, the vocoded speech channel was only

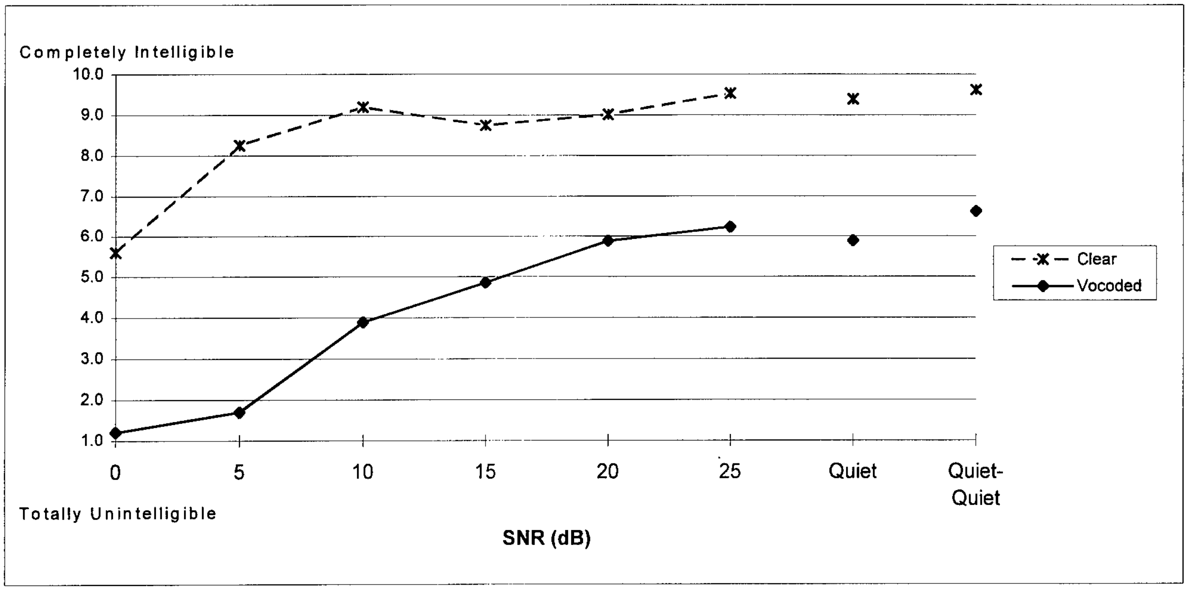


Figure 6 - Intelligibility of Contextual Sentences

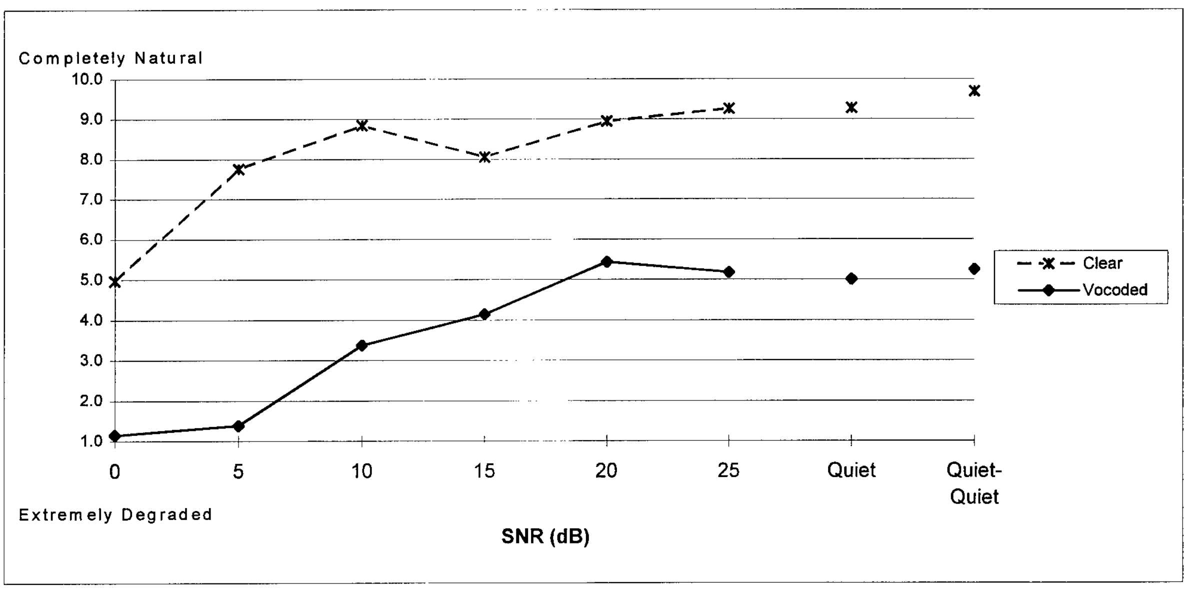


Figure 7 - Quality of Contextual Sentences

Talker Noise	Listener Noise	Clear Speech	Vocoded Speech
0dB SNR	Helicopter	5.6	1.2
5dB SNR	Helicopter	8.3	1.7
10dB SNR	Helicopter	9.2	3.9
15dB SNR	Helicopter	8.8	4.9
20dB SNR	Helicopter	9.0	5.9
25dB SNR	Helicopter	9.5	6.2
Quiet	Helicopter	9.4	5.9
Quiet-Quiet	Quiet	9.6	6.6

Table 4 - User Assessment of Intelligibility

Talker Noise	Listener Noise	Clear Speech	Vocoded Speech
0dB SNR	Helicopter	5.0	1.2
5dB SNR	Helicopter	7.8	1.4
10dB SNR	Helicopter	8.8	3.4
15dB SNR	Helicopter	8.1	4.1
20dB SNR	Helicopter	8.9	5.4
25dB SNR	Helicopter	9.3	5.2
Quiet	Helicopter	9.3	5.0
Quiet-Quiet	Quiet	9.7	5.3

Table 5 - User Assessment of Quality

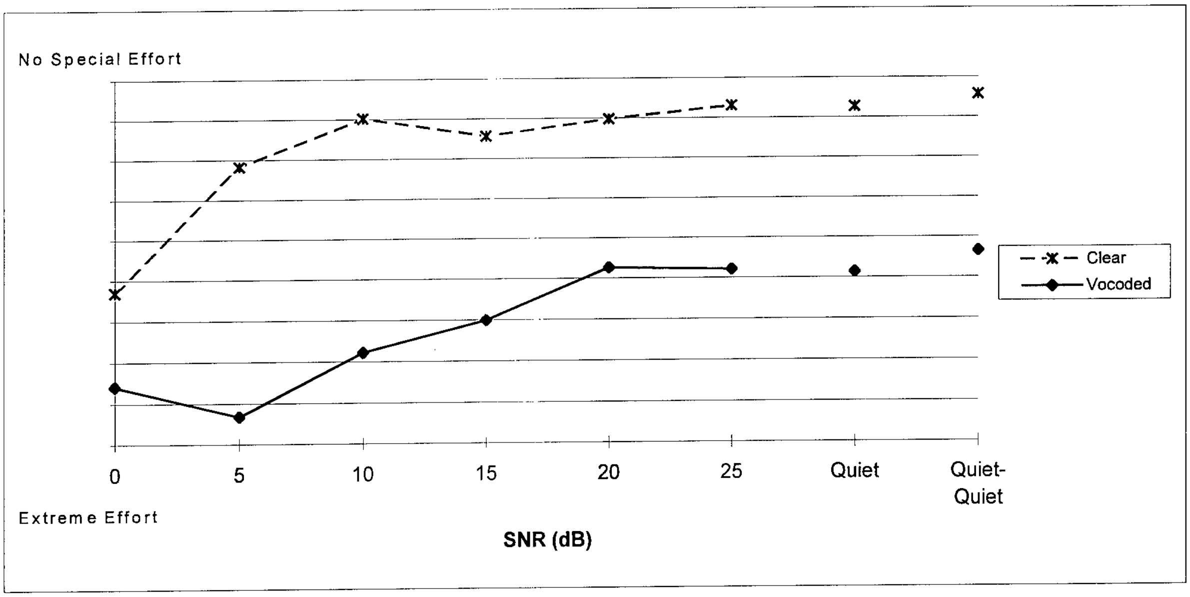


Figure 8 - Effort Required to Listen to Contextual Sentences

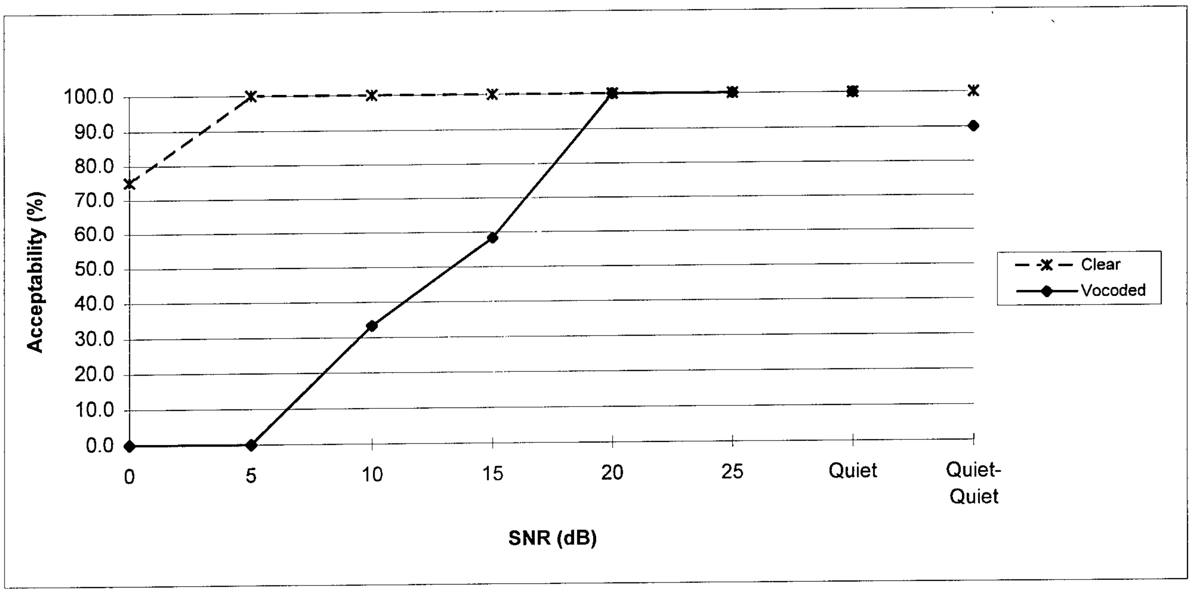


Figure 9 - Percentage of Aircrew Rating Contextual Sentences as Acceptable

Talker Noise	Listener Noise	Clear Speech	Vocoded Speech
0dB SNR	Helicopter	4.7	2.4
5dB SNR	Helicopter	7.8	1.7
10dB SNR	Helicopter	9.0	3.2
15dB SNR	Helicopter	8.6	4.0
20dB SNR	Helicopter	9.0	5.3
25dB SNR	Helicopter	9.3	5.2
Quiet	Helicopter	9.3	5.1
Quiet-Quiet	Quiet	9.6	5.7

Table 6 - User Assessment of Effort Required

Talker Noise	Listener Noise	Clear Speech	Vocoded Speech
0dB SNR	Helicopter	75.0%	0.0%
5dB SNR	Helicopter	100.0%	0.0%
10dB SNR	Helicopter	100.0%	33.3%
15dB SNR	Helicopter	100.0%	58.3%
20dB SNR	Helicopter	100.0%	100.0%
25dB SNR	Helicopter	100.0%	100.0%
Quiet	Helicopter	100.0%	100.0%
Quiet-Quiet	Quiet	100.0%	90.0%

Table 7 - User Assessment of Acceptability

rated as acceptable by 100% of aircrew when the talker SNR was greater than 15dB.

Below 20dB SNR, use of the pre-processors increased the percentage of aircrew rating the vocoded system as acceptable by up to 50% (i.e. an additional 6 of the 12 subjects rated the system as acceptable).

6. CONCLUSIONS

The pre-processors have demonstrated that improvements in STANAG 4198 vocoder performance in helicopter noise are possible at talker SNRs below 15dB. Increases in DRT score of up to 15% have been achieved. In addition, the user assessments of intelligibility, quality, effort required to listen and acceptability are all improved.

However, the user ratings of intelligibility, quality and effort required to listen are much lower for the vocoded channel than for the clear channel under all conditions (including quiet), even if a pre-processor is used.

7. REFERENCES

[1] NATO STANAG 4198, "Parameters and characteristics that must be common to ensure interoperability of 2400 bits/s linear encoded digital speech", February 1984.

[2] Papamichalis P.E., "Practical approaches to speech coding", New Jersey, Prentice-Hall Inc., 1987.

[3] Voiers W.D., "Evaluating processed speech using the Diagnostic Rhyme Test", Speech Technology, January/February 1983, pp 30-39.

[4] ANSI S3.5, "Method for the calculation of the Articulation Index", American National Standards Institute, 1969.

[5] Steeneken H.J.M. and Houtgast T., "A physical measure for speech-transmission quality", J. Acoust. Soc. Am., *67*, January 1980, pp 318-326.

[6] Interim Defence Standard 00-25, "Human factors for designers of equipment", Part 9:Voice Communication, Issue 1, 30 April 1991.

[7] Pratt R.L., Flindell I.H. and Belyavin A.J., "Assessing the intelligibility and acceptability of voice communications systems", RSRE Report 87003, June 1987.

[8] "Procedure for the operation of the Speech Intelligibility Facility", DRA/AS/SI/MMI/P9534/2.0, March 1996.

8. ACKNOWLEDGEMENTS

The UK MOD for funding the work described in this paper.

Paul Kellett (DRA) for his assistance in conducting the trials described in this paper.

All the subjects who participated in the DRT tests and to the aircrew who gave their time to take part in the user assessment trial.

ANNEX A - DIAGNOSTIC RHYME TEST VOCABULARY

VOICING			NASALITY			SUSTENTION		
Voiced	--	Unvoiced	Nasal	--	Oral	Sustained	--	Interrupted
veal	--	feel	meat	--	beat	vee	--	bee
bean	--	peen	need	--	deed	sheet	--	cheat
gin	--	chin	mitt	--	bit	vill	--	bill
dint	--	tint	nip	--	dip	thick	--	tick
zoo	--	sue	moot	--	boot	foo	--	pooh
dune	--	tune	news	--	dues	shoes	--	choose
voal	--	foal	moan	--	bone	those	--	doze
goat	--	coat	note	--	dote	though	--	dough
zed	--	said	mend	--	bend	then	--	den
dense	--	tense	neck	--	deck	fence	--	pence
vast	--	fast	mad	--	bad	than	--	dan
gaff	--	caff	nab	--	dab	shad	--	chad
vault	--	fault	moss	--	boss	thong	--	tong
daunt	--	taunt	gnaw	--	daw	shaw	--	caw
jock	--	chock	mom	--	bomb	von	--	bon
bond	--	pond	knock	--	dock	vox	--	box

SIBILATION			GRAVENESS			COMPACTNES		
Sibilated	--	Unsibilated	Grave	--	Acute	Compact	--	Diffuse
zee	--	thee	weed	--	reed	yield	--	wield
cheep	--	keep	peak	--	teak	key	--	tea
jilt	--	gilt	bid	--	did	hit	--	fit
sing	--	thing	fin	--	thin	gill	--	dill
juice	--	goose	moon	--	noon	coop	--	poop
chew	--	coo	pool	--	tool	you	--	rue
joe	--	go	bowl	--	dole	ghost	--	boast
sole	--	thole	fore	--	thor	show	--	so
jest	--	guest	met	--	net	keg	--	peg
chair	--	care	pent	--	tent	yen	--	wren
jab	--	gab	bank	--	dank	gat	--	bat
sank	--	thank	fad	--	thad	shag	--	sag
jaws	--	gauze	fought	--	thought	yawl	--	wall
saw	--	thaw	bong	--	dong	caught	--	taught
jot	--	got	wad	--	rod	hop	--	fop
chop	--	cop	pot	--	tot	got	--	dot

Development and performance of a cockpit control system operated by voice

H.J.M. Steeneken* and E.W. Pijpers**
* TNO Human Factors Research Institute, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
** National Aerospace Laboratory NLR, P.O. Box 90502, 1006 BM Amsterdam, The Netherlands

1. SUMMARY

The hands and eyes busy situation and high workload as is normally the case in a fighter cockpit require a natural means of communication between operator and system. Voice is an obvious means of communication, however the adverse cockpit conditions deteriorate the speech signal in such a way that automatic recognition of spoken commands is difficult.

In this study the performance of a voice controlled cockpit was investigated. An automatic speech recognizer was integrated in the control system of a F-16 simulator. The performance was evaluated during representative operational simulated flights. A similar study was performed by Prévôt and Onken (1995).

The results indicate that a flexible syntax of the commands is required. The recognition performance (75%) of the connected command strings was not good enough to accommodate the pilots.

From flight tests a representative spontaneous speech data base was collected for further improvements.

2. INTRODUCTION

The increasing complexity of aircraft systems coupled with the requirements to operate in all weather conditions at very low level creates a high workload for the pilot. The primary interest is to fly the aircraft which results into an eyes and hands busy situation. The control of other systems should not require too much of the pilot's attention. The use of speech for this purpose can be seen as a logic means of operation as long as the dialogue used for the control is simple and natural. However, the adverse environmental conditions in a cockpit, such as the high noise level and the microphone mounted inside an oxygen mask decrease the performance of a speech recognition system. This is especially the case if instead of artificial isolated words the more natural connected words are used. The design of a robust electro-acoustical input system and a logic vocabulary and syntax are therefore essential. In the present study a system was developed, based on a commercial recognizer.

Validation of the system was performed in the National Simulator Facility (NSF) which is based on a Mid-life Update F-16 cockpit.

3. AUTOMATIC SPEECH RECOGNITION IN A FAST-JET COCKPIT

In 1991 a project was started to study the use of speech recognition for cockpit control tasks. It was identified that a number of tasks are suitable for voice control. The study was divided into three phases:
(I) selection of the control tasks, feasibility, and selection of a commercial state-of-the-art recognizer,
(II) compilation of a vocabulary and command syntax structure based on the identified control tasks, and development and integration of the recognition system with a flight simulator,
(III) evaluation and data collection of the system during representative simulated operational flights.

Phase I

The tasks that were selected for voice control concern Data Entry, Display Management, Hands-On-Throttle-And-Stick (HOTAS), and so-called "Crew Assistant" control functions. In Fig. 1 an overview of the F-16 simulator cockpit and the selected controls is given.

The requirements for the recognizer were: connected word recognition (300 words), robust for background noise, short response time, and possibility of integration into a system. Based on these requirements only two candidates were available in 1991. It was identified that for a fair performance of the recognizer, given the poor speech input conditions (oxygen mask, noise, speech level variations), additional signal processing was required.

Phase II

The two main goals of phase II were development of vocabulary and syntax, and development and assessment of the recognizer integration.

A vocabulary consisting of 281 functional control words was compiled. It includes only a small amount of synonyms (such as "nav" and "nev" for the same control action) but also similar sounding words such as U_H_F and V_H_F.

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

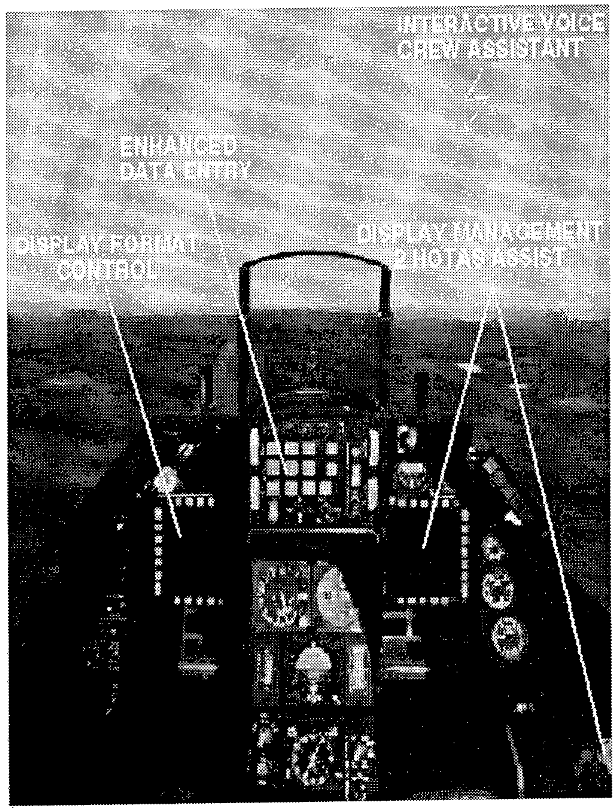


Fig. 1. Set-up of the voice controlled F-16 simulator cockpit.

The syntax required for the selected functions has 300 nodes and over 4000 node-to-node connections.

The vocabulary words and syntax are based on standard pilot vocabularies, and they were refined during "demonstration" stages where words with a low recognition or high insertion rate were deleted; the words were further checked for maximum discrimination.

For example some very frequently used words are: *counter, checked, display, master, chaff, radar, steer point, enter, gimme, up, select.*

The syntax design differentiates between so-called:
(1) selection commands: generally, consisting of two words e.g. "*display F_C_R*", and
(2) data entry sequences consisting of 5–10 words e.g. "switching U_H_F 374.9 enter".
Apart from the standard "*return*" commands, some functions are readily accessible to the pilot by specific design features. Frequently used command strings are:

master nev,
select steer point auto,
radar declutter,
display map,
switching U_H_F tactical.

During phase II also the recognizer system was evaluated. The application required input of the speech through a microphone placed in an oxygen mask. The cockpit noise level can be up to 105 dBA and the speech level may vary due to variation of the speaker's vocal effort. In order to obtain a speech input signal for the recognizer with a fair quality an optimal design of the electro-acoustical interface was required.

A specific noise cancelling microphone was used. The speech level was controlled by an automatic gain control amplifier (AGC) which stabilized the signal level. As no representative speech data were available a data base was recorded in the laboratory with five speakers. Both isolated words and connected word strings were used. The speakers equipped with helmet and oxygen mask were placed in a high noise room where a representative diffuse sound field up to 110 dBA could be obtained. Digital recordings with and without background noise were made. During the recordings the speaker was supplied with a side-tone in order to stabilize his vocal effort. This is identical with the situation in the cockpit.

For the evaluation of the recognizer a specific test-bed was used which was controlled by a workstation. Initialisation, parameter setting, training and testing

were performed automatically. In this way the effect of various parameters was studied. This includes: AGC setting, training method, speaker dependency, word mode (isolated, sentence), and syntax. In Table I the recognition performance is given for three speakers and four noise conditions.

The training for all conditions was the same and performed at a noise level of 100 dBA. The performance is expressed by the ratio correctly recognized and by an accuracy measure which includes also insertions and deletions (Hunt, 1990). The scores are high but it should be noted that the speech data bases used for the experiments were based on read speech rather than on spontaneous speech. This also implies that the pilots could not make syntax errors as the text to be read was prompted on a screen. Table I gives the final results of all tests after a long optimalization period.

The system was integrated into the National Simulator Facility which is available at the National Aerospace Laboratory. The voice input system offers direct control of the cockpit control systems indicated in Fig. 1.

Table I. Recognition performance for the cockpit control sentences and words (accuracy for words, percentage correct for sentences) for three pilots and read speech. Four noise conditions were included: no noise and with noise (95, 100, and 105 dBA). The training for both conditions was *with* noise (100 dBA).

Speaker		No noise		95 dBA		100 dBA		105 dBA	
		words acc.	sent corr.	words acc.	sent corr.	words acc.	sent corr.	words acc.	sent corr.
Pilots	P1	0.90	0.86	0.98	0.97	0.99	0.97	0.98	0.95
	P2	0.97	0.95	0.98	0.97	0.99	0.98	0.97	0.93
	P3	0.99	0.99	0.98	0.97	0.99	0.99	0.97	0.94
mean		**0.95**	**0.93**	**0.98**	**0.97**	**0.99**	**0.98**	**0.97**	**0.94**
se		**0.03**	**0.04**	**0.00**	**0.00**	**0.00**	**0.01**	**0.00**	**0.01**

Phase III

The goal of the project on voice control of cockpit systems is to perform a realistic experiment and to obtain subjective pilot responses and objective performance measures.

Real-time recognition and control of systems was performed during realistic flights (sorties), with an average duration of 70 min. The pilot could operate a number of systems either by voice or manually. In total 17 sorties were performed with voice control. The total flight time was 18 hrs. The total amount of speech utterances for the control task was 134 min. Three pilots participated in the experiments. All these pilots were familiar with the standard F-16 cockpit and flight procedures.

During the tests the output of the recognizer was stored into a logfile together with the speech signal and the PTT-actions. In the laboratory the speech signal was transcribed (annotated) orthographically. Hence, from all spoken utterances the written version was available. By comparison of the recognizer response (logfile) and the annotation file the performance can be obtained.

An automatic scoring program gives the words correct, deletions and insertions. From this the accuracy measure can be calculated.

The speech signals were also recorded for later analysis and to repeat the experiment under laboratory conditions. This is relevant for conditions where syntax errors, false PTT-triggers (push-to-talk), hesitations, etc. define the performance of the presently used recognizer. Finally with the collected speech material a calibrated data base was compiled.

The mean accuracy measure thus obtained for each pilot is given in Table II (header NSF). The accuracy is very speaker dependent, pilot #2 gives the highest scores (acc. 81%) and the mean of all three pilots is 69%. Errors may be introduced by a poor recognition performance of the system but also by the speakers (e.g. syntax errors, uttering out-of-vocabulary words (OOV's), and incorrect PTT-actions).

Table II. Mean performance (accuracy %) for the simulator experiment and the laboratory replay with the annotated speech material for three pilots.

	NSF	Lab.	No syntax
pilot #2	81	92	84
pilot #3	66	70	40
pilot #4	60	64	58
all pilots	69	75	61

In order to evaluate these control errors all the speech material was transcribed to computer files, annotated, and corrected for operation errors of the pilots. The recognizer performance was measured in the laboratory test-bed with this corrected database. The results are also given in Table II. An improvement of the scores of 6% was obtained.

Analysis of the words used in the command string by the pilots showed that from the original 281 word vocabulary only 175 words were used. There were also 42 additional words used which were not in the vocabulary. In total 12231 words concentrated in 5825 utterances were analysed. It was found that with 65 words a 90% coverage is obtained for all tested conditions. With these 65 words the experiments were repeated without making use of a syntax. The corresponding performance of the recognizer is given in Table II column "no syntax". A degradation of 14% with respect to the laboratory test is obtained. For the condition with the syntax the average perplexity (words open for recognition) amounts to 13.5, and consequently the perplexity without syntax was 65.

4. CONCLUSION

The voice input system developed for control of systems in the F-16 cockpit was successfully integrated in a flight simulator. During the system demonstrations the speech recognition performance was inconsistent for each pilot and variable over time. The pilot judgement ranged from unsatisfactory to being surprised at the state-of-the-art and the relatively good performance. A detailed analysis of the performance and of the pilot's comments is still in progress. The initial results indicate a requirement for more flexibility in the syntax design, more consistent recognizer performance and a re-allocation of cockpit control functions to voice control with emphasis on new "crew assistant" functions.

The speech recognizer should be able to handle spontaneous speech in order to improve the performance.

From the simulator trials a representative calibrated data base was obtained which is useful for assessment of present speech technology systems.

5. ACKNOWLEDGEMENTS

The authors are indebted to D.A. van Leeuwen and J.J. Kriekaard of TNO-HFRI and to J. Eertink and P. van Zutphen of NLR for their valuable contribution in the development of the system and the performance and analysis of the tests.

6. REFERENCES

Hunt, M.J. (**1990**). "Figures of Merit for Assessing Connected-Word Recognisers", *Speech Communication*, Vol 9, pp. 329-336.

Prévôt, T., and Onken, R., (**1995**). "In-flight evaluation of CASSY: A system providing intelligent on-board pilot assistance", Air Traffic Control Quarterly, Vol.3(3) 183-204.

Steeneken, H.J.M. (**1992**). "Quality evaluation of speech processing systems", Chapter 5 in *Digital Speech Coding: Speech coding, Synthesis and Recognition*, edited by Nejat Ince, (Kluwer Norwell USA), 127-160.

VOICE RECOGNITION IN ADVERSE AIRCRAFT COCKPIT ENVIRONMENTS

A J South
Systems Integration Department
Probert Building
DERA Farnborough
Ively Road
Farnborough
Hants. GU14 0LX
UK

SUMMARY

A speech recognition system has been flown in a two seat Tornado strike aircraft and assessments made of the recognition accuracy under normal, terrain following and 4G turning flight.

Word accuracies averaged some 96% under normal flight, and 95% under terrain following. During 4G turns the recognition levels dropped to around 80%.

Subsequent speech recordings made on the centrifuge at the RAF School of Aviation Medicine consisted of lists of digit strings and typical Direct Voice Input command phrases. Recordings were made at up to 8G, using four different levels of anti-G protection. The subjects were five male RAF personnel, and one female.

The digit string lists were used to test a speaker-dependent whole word speech recogniser at up to 6G.

The results will be presented for protection using standard and full coverage anti-G garments, and with the use of positive pressure breathing. Possible solutions to the lower accuracy rates at higher G levels and with pressure breathing are discussed.

1. INTRODUCTION

The potential benefits of using automatic speech recognition in the military cockpit have long been recognised (Ref. 1). The modern cockpit is a high workload environment, especially for single seat aircraft. The many systems and equipment that have been added to aircraft over the past twenty to thirty years have improved their capability enormously. However, each one adds monitoring and control functions for the pilot to carry out, in addition to his primary task, which must always be to fly the aircraft.

The pilot's control over an aircraft relies primarily on two channels of information: eyes and hands. The pilot assesses the situation mainly through his eyes and commands the aircraft through his hands. There are situations where both of these channels may be busy, and yet the mission demands that the pilot makes extra control inputs, such as re-tuning the radio. In this case, the voice would be a natural channel for the pilot to use.

This potential was recognised over fifteen years ago (Ref. 2) and, over the years, several flight trials and other experiments have been carried out at the DRA. In 1984, a connected speech recogniser (the Marconi SR128) was flown in a BAC 111 civil airliner, based at Bedford (Ref. 3). Over a period of time, the recognition system was developed as an input to the radio, navigation and flight management systems. As well as carrying out trials on the performance of the recogniser, two of the pilots (who achieved high recognition accuracy) also used the system as part of their cockpit interface while carrying out other trials on the aircraft. The same recogniser was also flown on a Wessex helicopter at Bedford, and a Buccaneer strike aircraft at Farnborough.

Following these trials, a specification was written for a flightworthy speech recogniser, using the latest algorithms and with capacity for a vocabulary of one thousand words. The ASR1000 was delivered early in 1989, and tested extensively on recordings made on board a Tornado that same year. At about this time, a new algorithm (the IMELDA transform) was being developed which promised to bring about a substantial reduction in the error rate (Ref. 4). It also had the capability of being optimised for particular microphones and background noises. A transform optimised for Tornado was developed and fitted to the ASR1000 in 1992. The following year, more Tornado flight trials were carried out, which are reported below (Section 2).

While automatic speech recognition is rapidly gaining acceptance for commercial uses, in offices and for telephone enquiry systems, there is still no operational use of such systems in military aircraft. The high recognition accuracy requirement and the noisy, stressful environment make the development of a practical system very difficult. The noise, vibration, G force and other stresses of the military cockpit environment all affect the way in which the crew speak. All current speech recognition systems work by comparing the incoming speech with previously stored examples of the words or phonemes. If a high accuracy is required, the stored word templates will be particular to the user. It follows that anything that has the effect of altering the quality of the user's speech is likely to degrade the performance of the recogniser.

For some while now, progress in automatic speech recognition has been largely data driven. Performance has been improved by using larger and larger training and testing databases, and hastened by the concurrent increases in available computing power. It may seem

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

surprising to the non-specialist that relatively little use has been made of phonology, the study of the sound systems of language. In consideration of the particular changes that may be induced in voice quality due to the physical stresses of flight in military aircraft, it seems likely that a detailed study of these effects will help in the development of robust recognition algorithms which will allow the full benefits of this technology to be realised.

Section 3 of this paper describes a series of recordings made on the centrifuge at the RAF School of Aviation Medicine at Farnborough. The aim of this trial was to study the effect of G force on speech production and recogniser performance, in isolation from the other stressors that are present in flight. As well as describing the recording conditions and materials, the results of preliminary experiments on the recognition of digit strings are presented here.

Section 4 discusses these results and considers some possibilities for improving recognition performance under these difficult conditions.

2. TORNADO TRIALS

2.1 The Installation

While the ASR1000 speech recogniser was being developed, the air fleet of the DRA was examined to find the most appropriate aircraft for flight trials to demonstrate the use of speech recognition in the cockpit. The Tornado was chosen, as being the most representative of modern fast-jet aircraft, and bearing in mind that speech recognition was to be included in the European Fighter Aircraft.

One of the main pieces of equipment used by the navigator in the rear seat of the aircraft is the Television Tabular display, known as TV-TABS for short. This is used to display the route, enter navigation data, and for various related functions during the attack phase of the mission. As panel space is always limited in military cockpits, the TV-TABS has a small keyboard underneath the display (see Fig. 1). There are dedicated keys for the main display modes, and a row of soft keys below the screen, the menu for which appears along the bottom edge of the screen. Other keys call up menus for letters and numbers, and there is a two-way toggle switch to move a cursor along the read-out line, which appears just above the legends for the soft keys.

The TV-TABS implements about forty functions related to navigation and attack. With such a limited keyboard, there is inevitably a complex structure of nested menus, requiring many keystrokes to carry out most functions. This is not an easy interface to use and is not popular with the aircrew. On examining the functions and menu structures, it was found that a vocabulary of about one hundred words would suffice for voice input. The syntax could be arranged to allow direct access to functions without having to go through several levels of menus. In addition, it was possible to make a simple electrical connection into the system. The interface could also allow reversion to normal operation with a single switch, an important safety consideration.

The main aim of the trial was to allow the navigator to operate the TV-TABS by voice, and make comparisons of the time taken for various operations between voice and keyboard. Unfortunately, there were some problems with the software in the interface that could not be solved in time for the flight trials, so live operation of the TV-TABS by voice was not possible. However, it was possible for the navigator to read lists of command phrases, and to see the response of the recogniser. All his speech and the recogniser's responses were recorded, as well as cockpit noise and many of the aircraft's flight parameters.

The final vocabulary used consisted of 99 words, including the digits and letters. The recogniser's syntax mimicked the keyboard sequences, but also allowed direct access to functions without having to go through all levels of the menu. The syntax branching factor was about 15. The recogniser's word models were trained from isolated utterances only, using 20 examples for the digits, 15 for letters and 10 for the other words.

During the flights, the navigator was prompted to read lists of command phrases from a special display that was fixed to the top of his instrument panel. The recogniser's responses were also displayed here. A total of fifty phrases were included, with lengths ranging from two words to ten. The short phrases mostly concerned changes of display mode. Longer phrases, for entry of waypoints, for example, contained digit strings as well as some other words.

In addition to the command phrases, some lists of isolated digits and digit triples were included. The digits form the most difficult subset of words to recognise (in English, at least), and so make the most sensitive test of the recogniser's performance. In practice, of course, much of the data that may be input by voice would consist of digits and so it is important to determine the recognition performance for this subset of the vocabulary.

2.2 Flight Conditions

Several different flight conditions will normally be encountered in the course of an operational sortie. High level transit may be followed by very low level ingress to the target. High G manoeuvres may be required. In the ideal case, the crew should be able to use the speech recogniser under all conditions. Three flight conditions were included in the trials, to cover the main range of possibilities:

- Straight and level.
- Simulated Terrain Avoidance.
- Continuous 4G turns.

All conditions were intended to be flown at 4,000 ft. although sometimes weather conditions or air traffic control dictated otherwise. The speed was 420 knots for all conditions. The terrain avoidance condition had to be

simulated because most of the flights took place during the winter when flight at 250 ft. through mountainous territory was not usually possible. Although flying at 4,000 ft. or more, the pilot was instructed to manoeuvre the aircraft as if terrain following, while the navigator read the lists.

Measurements of the cockpit noise showed average sound pressure levels of 100 dB in straight and level flight, rising to 112 dB in the 4G condition. The Tornado has a relatively quiet cockpit; other comparable aircraft have noise levels about 10 dB higher.

Over a period of about eight months, 19 sorties were carried out using six subjects. A total of about 15,000 words of speech were recorded in the air. Additional recordings of about 18,000 words were made in the laboratory for training the speech recogniser, and for other related experiments.

2.3 Results

The performance of a speech recogniser is usually expressed in terms of word accuracy, defined as:

$$\text{Acc.} = \frac{\text{No. of words spoken - No. of errors}}{\text{No. of words spoken}} \times 100\ \%$$

Errors may be substitution of one word for another, no response to a word (deletion), or insertion of an extra word.

Table 1 shows the word accuracies achieved for isolated digits under the three flight conditions. As well as the average across all speakers, the results for the best and worst speakers are shown in each case, in order to show the variation of performance between speakers. These are not necessarily the same speaker in all flight conditions. Table 2 shows the results for the lists of digit triples. Recognition of connected speech is more difficult than isolated words as a rule, but is more realistic from the user's point of view.

Calculation of the word accuracy is more complicated when whole phrases are being recognised with a syntax operating. If one error occurs, the wrong branch of the syntax may be followed with the likelihood that all subsequent words in the phrase will be mis-recognised. In order to generate a fair recognition score for the recogniser, only the first error is counted; subsequent words are ignored for both the error count and the count of words spoken. An additional difference occurs because, when matching the recogniser output with the phrase spoken using a dynamic programming algorithm (Ref. 5), a word substitution is impossible to distinguish from a deletion-insertion pair, unless a careful examination is made of the start and finish times of the words. For this reason, deletion and insertion errors are weighted by a factor of 0.5 in the calculation of word accuracy for the phrases.

Table 3 shows the word accuracies achieved on the later sorties of the trial. Changes made to the syntax during the course of the trials make the results from the earlier flights incompatible.

It can be seen from the results that for all three lists, the results achieved by the best speaker in straight and level flight were over 98%. This is the level at which pilot acceptability seems fairly certain. The average accuracy for the command phrases of 95.7% is perhaps not acceptable, but it should be remembered that the longer command phrases (five words or more) contain strings of digits. The accuracy for connected digits (Table 2) is only 92.9%. If only the short phrases are considered (i.e. those not containing digits), an average word accuracy of 97.8% is obtained.

In all cases, the results show a considerable decrease in accuracy under the 4G condition. This is especially true for those subjects who had least experience of fast-jet aircraft. It is only to be expected that such a high physical stress will alter the speaker's voice considerably, leading to a poor recognition accuracy. However, it is encouraging that one subject, who had over 3,000 flying hours on fast-jets, achieved nearly 92% accuracy under 4G, only about 3% less than his straight and level result.

2.4 Conclusion

These results show that present speech recognition technology is capable of achieving adequate performance for some speakers under the least stressful flight conditions, but improvements in robustness must be made before the technology will be good enough for operational use.

It was decided that studies should be made of the effects of the military cockpit environment on speech production. By examining each in isolation from the others, it is hoped that ways may be found to improve recognition algorithms to make them more robust to these effects.

3. CENTRIFUGE TRIALS

3.1 The Recordings

There were two main objects of this experiment: firstly, to make recordings that could be used to assess the performance of speech recognition equipment, and secondly to study the way in which the acoustic-phonetic parameters of speech change when the speaker is subjected to high G levels. This paper reports on the data collection and recognition experiments only.

The centrifuge at the RAF School of Aviation Medicine at Farnborough is sufficiently well known not to need detailed description here. The gondola was fitted with a personal computer monitor to prompt the subject, and a cockpit communication system (CCS) station box to which the subject's headset lead was connected. A meter showing the speech level was also fitted to help the subject to maintain a constant vocal effort. There was also a hand-held switch for the subject to press while speaking, the Press-To-Recognise or PTR switch. This is used in a similar way to the Press-To-Talk switch for the radio. In a real cockpit, it would allow the recogniser to operate only on speech intended for it. In

this installation, the PTR also controls the prompting computer.

The prompting computer, digital audio tape recorder and CCS master were installed at one of the operator stations near the axis. The CCS was wired into the normal intercom so that the centrifuge operator and medical staff could communicate with the subject as normal.

For this experiment it was decided to use a vocabulary and syntax that had been proposed for the EF2000 project. This vocabulary was subsequently superseded for the application, but is reasonably representative of the cockpit control task. There are 104 words, including digits and some (but not all) letters, and the mean branching factor of the syntax is about 8. A list of 25 command phrases was generated from the syntax. Where a command would normally include long digit strings, as in geographical co-ordinates for example, these were shortened to one or two digits. A separate list of 5-digit strings was included in the recordings to characterise performance on connected digits. Strings of five digits were used, rather than the standard lists of triples, in order to reduce the time taken to read each list.

Separate recordings were made in the laboratory from each speaker, in order to collect material for training the recogniser. These included mainly isolated word lists, but also a list of digit strings and a list of command phrases.

A short list of eleven phonetically rich sentences, taken from the SCRIBE project (Ref. 6), was also included in the training recordings and the first set of centrifuge recordings.

The first five subjects were all RAF personnel employed at the School of Aviation Medicine. Two were fast-jet pilots and three were doctors. All were experienced at riding the centrifuge. The recordings were made in two sessions, separated by several months. During the second session one of the subjects was not available, but a substitute was found. This subject was a female, who had considerable experience on the centrifuge, but had not been a subject for about two years.

3.2 Anti-G Protection Conditions

In consultation with the centrifuge staff, it was decided to make recordings over a range of G levels, with different types of anti-G protection. In particular, it was desired to make comparisons between the current standard anti-G trousers and the full coverage type under consideration for future high performance aircraft such as EF2000. This aircraft is also intended to use positive pressure breathing, which can be expected to have a significant effect on speech production. At the time of the recordings (1994), the aircrew equipment fit for the aircraft had not been finally decided, but the best available estimate was used. The following conditions were used for the recordings.

- No protection 1,2,3 G
- Standard anti-G trousers 3,4 G
- Full coverage anti-G trousers (FAGTs) 3,4,5 G
- FAGTs plus pressure breathing 4,5,6 G

A short list of five command phrases was also recorded at up to 8G with the latter two protection conditions.

In addition to the anti-G protection, subjects wore the standard flying kit of overalls, life jacket, flying helmet and oxygen mask.

The standard lists took between 90 and 120 seconds to read. The prompting and recording system allowed the subjects to pause in the middle of a list if they wished, but none of the subjects took advantage of this. In general, subjects took rest periods of a minute or two between lists. In a few cases they read two or three lists without a break at the low G levels. Each subject had three sessions on the centrifuge, the first two protection conditions being combined into one session. The total time for each session was about 45 minutes. For each condition, two subjects recorded the lists starting from the highest G level and working down, while the other three started with the lowest level and worked up. This was done to try to balance out fatigue effects.

The recordings were digitised and annotated in the format prescribed by the ESPRIT Speech Assessment Methods project (Ref. 8). The data is available on CD-ROM.

3.3 Speech Levels under G

The first measurements made on the recordings were of the speech level, related to loudness. The first five command phrases of each speaker under each condition were measured using the program SAM_SLM (Ref. 7). These figures were related to absolute sound pressure levels by means of the 94 dB calibration tone recorded on the tape at the start of each session.

The absolute speech levels of the subjects varied between 100 dB and 114 dB in the 1 G condition, but what is of interest is how much the speech level changes as the G level is increased. Figure 2 shows the changes relative to the 1 G level, averaged across all speakers. Despite large differences between individual responses, there is a fairly consistent increase of 2 dB per G on average. This could be important for the use of a speech recogniser, as some types are quite sensitive to the level of the speech input. Normal speech may vary over a range of about +/- 6 dB from the mean; adding another 8 to 10 dB could have serious implications for the performance of most recognisers.

3.4 Recognition Results

To date, recognition tests have been carried out only on the digit lists. Word models were trained from twenty isolated utterances of each word, recorded in the laboratory. (One subject, AP, was not available for all of the training recordings, so only ten utterances were used for his models.)

Recognition tests were carried out automatically using the SAMPAC software developed under the ESPRIT

SAM project, but modified at Farnborough to make it more suitable for testing a recogniser that operates with a PTR switch. One list of digit strings was tested for each subject at each combination of protection and G level.

The one female subject gave results that were quite dissimilar from the other subjects. There could be many reasons for this, including the fact that she had not ridden on the centrifuge for about two years. While her results may give pointers to areas that should be researched in future, particularly as female aircrew become more common, it was thought wise to exclude them from the general analysis.

Figure 3 shows the recognition accuracy (calculated as described in section 2.3) averaged across the five male subjects for all protection conditions. The result for the test on the digit string list recorded with the training data is also included. A statistical test showed that there was no significant difference between the results on the laboratory recording (96.5%) and the 1G recording on the centrifuge (95.5%), thus confirming that there were no significant differences in the audio recording system.

The results show a steady, but not steep, decline in accuracy from 1G to 5 G for the first three protection conditions. There is no apparent difference between the various types of anti-G protection trousers, at least at 3 G and 4 G. The average at 5 G with FAGTs was 89.8%.

Comparison with the results from the Tornado trials in Table 2 suggests that the G level alone was not the only factor in the low figure achieved under 4 G in flight. However, such a comparison must be treated with caution because of the small number of subjects available for these experiments. There was also a big difference in the experience of the subjects; all of the centrifuge subjects were very experienced whereas three of the Tornado subjects had little experience of high G levels.

By way of illustration, Figures 4 and 5 show the best and worst subjects. Subject AP achieved not less than 97.6% at any G level in the first three protection conditions. He was the subject with the most experience on the centrifuge. ML, on the other hand, was the youngest and least experienced, but may also have been the least consistent speaker anyway. His result on the training data is only 88.8%, and in fact there appears to be little degradation as the G level increases

The inclusion of pressure breathing has clearly made a large difference to the results. At 4 G, the accuracy is down to 86.5% compared to 94.1% without pressure breathing. At higher levels, the recognition rate falls rapidly. This is not surprising, given that the excess pressure will tend to expand the vocal tract, thus changing its resonant frequencies. It also appears, to judge by the sound of many utterances, that the pressure makes closure or approximation of the articulators in stops and fricatives more difficult.

A further more serious problem also became apparent with two subjects. At higher G levels, air could leak around the edges of the oxygen mask, causing wailing or rasping noises. At times these noises were of similar loudness to the subject's speech. These two subjects both had high numbers of deletion errors in the pressure breathing condition, which is to be expected in a high level of background noise.

4. DISCUSSION

On average, the results obtained on the Tornado are, we believe, not quite good enough for operational use. The subjects were also asked to make subjective judgements of the performance of the recogniser. In over 60% of cases, the ratings were "satisfactory" or better. This figure should be treated with caution, because the speakers were only reading lists and not using the recogniser to carry out a real function. They were not required to correct errors to achieve a fully correct phrase. On the other hand, the additional motivation arising from using the recogniser for a real task would probably result in the users learning how to get the best results, with a subsequent improvement in recognition accuracy.

The recognition accuracy reported here is somewhat less than other similar airborne trials have achieved (see for example, Ref. 8). This may be partly due to the more complex syntax, with a branching factor of 15 compared with about 8 or less for other comparable trials. Differences in the flying experience of the subjects also appeared to be significant, although the numbers of subjects used was too small for a proper statistical assessment.

The reduction in accuracy under the 4 G condition demonstrates the extent to which the adverse environment of the military cockpit affects speech. In this case, it is not clear from the flight trial results whether the effect is due to the increased G or the higher level of noise in the cockpit. The Tornado has an unusually quiet cockpit for a fast-jet; in the straight and level condition, the noise is typically 100 dB SPL. This rises to 112 dB in a 4G turn. Although most speakers also increase their vocal effort under these conditions, this is not usually enough to compensate and the speech-to-noise ratio decreases.

The results of the centrifuge trial show that, for G-experienced speakers at least, the effect of up to 5 G is relatively small. It is unlikely that speech recognition would be required to operate at higher G levels in practice.

The use of positive pressure breathing is potentially of great use in allowing pilots of agile combat aircraft to function under high G levels, but it necessarily has a large effect on the vocal apparatus. The results shown here demonstrate that present speech recognition technology is unlikely to achieve an adequate performance under these conditions. Of all the environmental factors to be encountered in military aircraft cockpits, this is likely to be the most difficult to overcome.

These experiments have also shown that the fit of the oxygen mask could be a crucial factor in the performance of speech recognition equipment under high G with positive pressure breathing. Noises generated by leakage could also affect radio communications.

It is apparent from the above, that some improvements are still required before the full benefit of speech recognition technology can be realised in military aircraft cockpits. Some advances may arise from the development of speaker-independent recognition algorithms. It seems reasonable to suppose that a system that is robust to variations between speakers will also be robust to variations in one speaker resulting from environmental conditions. Support for this idea has been obtained in a small scale experiment carried out on the Tornado recordings. Compared with the speaker-dependent models, the performance of speaker-independent models was 1.5% worse in the straight and level condition, but 5% better in the 4G condition. This improvement resulted mainly from the accuracy on the poorest speakers being improved to a similar level to that of the others. The use of speaker-independent models would also save on the time taken to train and build models for each pilot, and the loading of the data to the aircraft before a sortie.

Speaker-adaptive algorithms can also be used. As little as a few seconds of speech can be enough to make a useful improvement in performance, but it is doubtful if the adaptation could occur rapidly enough to track changes in flying conditions.

In the longer term, it is felt that real progress in this area can only be made on the basis of a better understanding of the effects of the environmental stressors of the military cockpit on speech production. While it is widely recognised that individual responses to "stress" vary considerably, the effects of purely physical stressors such as G force, vibration, and positive pressure breathing are expected to be more consistent. Some studies of these effects have already been made (Refs. 9, 10, for example), but further work is needed to provide a full description on which techniques for robust speech recognition for military applications can be based.

5. CONCLUSIONS

The performance of an automatic speech recogniser in a fast-jet cockpit has been measured. Under the least stressful flight conditions and for some speakers, the recognition accuracy is high enough for operational use. The average performance is still 2-3% short of the level believed to be required. Under the stressful condition of continuous 4G turns, the performance decreases considerably on average, but only a few percent for the most experienced aircrew.

A considerable body of speech recordings has been made on a centrifuge, at up to 8 G. The speech recogniser has been tested on digit string lists recorded at up to 6 G, with relatively little performance loss, except when positive pressure breathing is introduced. This resulted not only from the effect of the pressure on the vocal tract, but also from noises introduced by leakage around the edge of the oxygen mask.

Techniques have been discussed which show promise of improving the performance of speech recognisers in the military cockpit, but it is considered that major advances can only be based on a sound understanding of the effects of the airborne environment on speech.

References

1. Becket, P., "Voice Control of Cockpit Systems", Proc. AGARD Symposium on Information Management and Decision Making in Advanced Airborne Weapon Systems, Toronto, April 1986

2. Smith, K., and Bowen, W. J. C., "The Development of Speech Recognition for Military Fast-Jet Aircraft." Proc. Military Speech Technology, October 1986.

3. Cooke, N., "RAE Bedford's Experience Using Direct Voice Input (DVI) in the Cockpit", Proc. Voice Systems Worldwide 1990, London, May 1990

4. Hunt, M. J., and Lefèbvre, C., "A Comparison of Several Acoustic Representations for Speech Recognition with Degraded and Undegraded Speech", Proc. ICASSP '89. 1989 pp262-265

5. Hunt, M. J., "Figures of Merit for Assessing Connected-Word Recognisers", Speech Communication 9 (1990) pp329-336.

6. "'SCRIBE' - Spoken Corpus Recordings in British English, A one year pilot project funded under the joint IED/DTI programme. August 1989.

7. "User Guide to Input Assessment", University College, London, SAM-UCL-G005, ESPRIT project 2589.

8. Thomas, J., "Advances in Speech Recognition" New Directions, Circuit 9, Spring 1993.

9. Bond, Z. S., Moore, T. J., Gable, B, "Acoustic-Phonetic Characteristics of Speech Produced in Noise and While Wearing an Oxygen Mask", J. Acoust. Soc. Am. 85 (2) February 1989.

10. Goldsmith, M., and Taylor, M. R., "Effects of Whole-Body Vibration on Speech", Proc. Institute of Acoustics, Vol 12: Part 10. 1990

	%Word Accuracy		
	Worst	**Best**	**Average**
Straight and Level	91.2	100.0	95.2
Terrain Avoidance	84.6	98.3	92.3
4G Turns	66.0	94.0	85.9

Table 1. Word Accuracy for Isolated Digits

	%Word Accuracy		
	Worst Speaker	**Best Speaker**	**Average**
Straight and Level	86.7	98.7	92.9
Terrain Avoidance	84.2	95.4	89.4
4G Turns	68.0	90.7	80.4

Table 2. Word Accuracy for Connected Digits

	%Word Accuracy		
	Worst Speaker	**Best Speaker**	**Average**
Straight and Level	93.7	99.6	95.7
Terrain Avoidance	94.6	97.9	95.5
4G Turns	80.0	91.9	85.3

Table 3. Word Accuracy for Command Phrases.

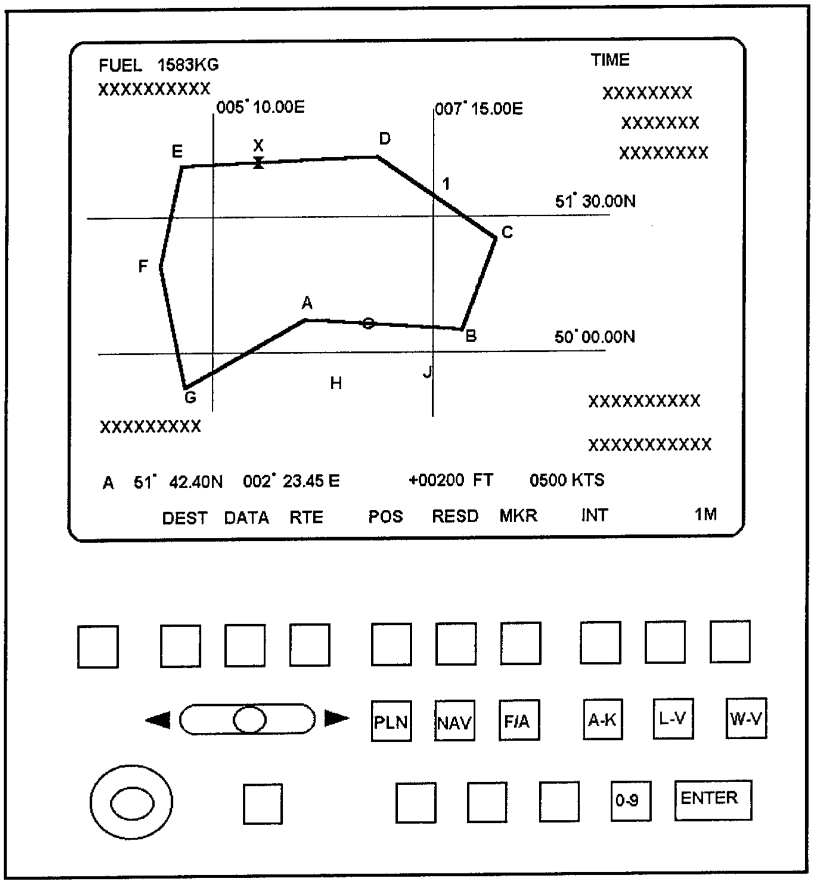


Figure 1. TV-TABS Display and Keyboard

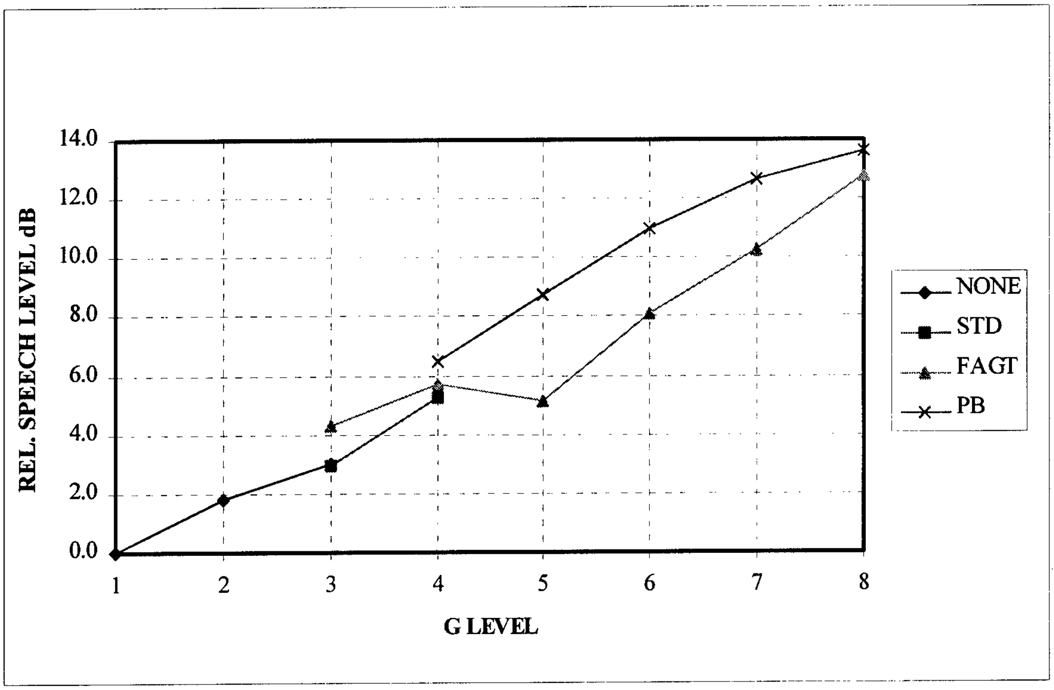


Figure 2. Change of Speech Level vs. G

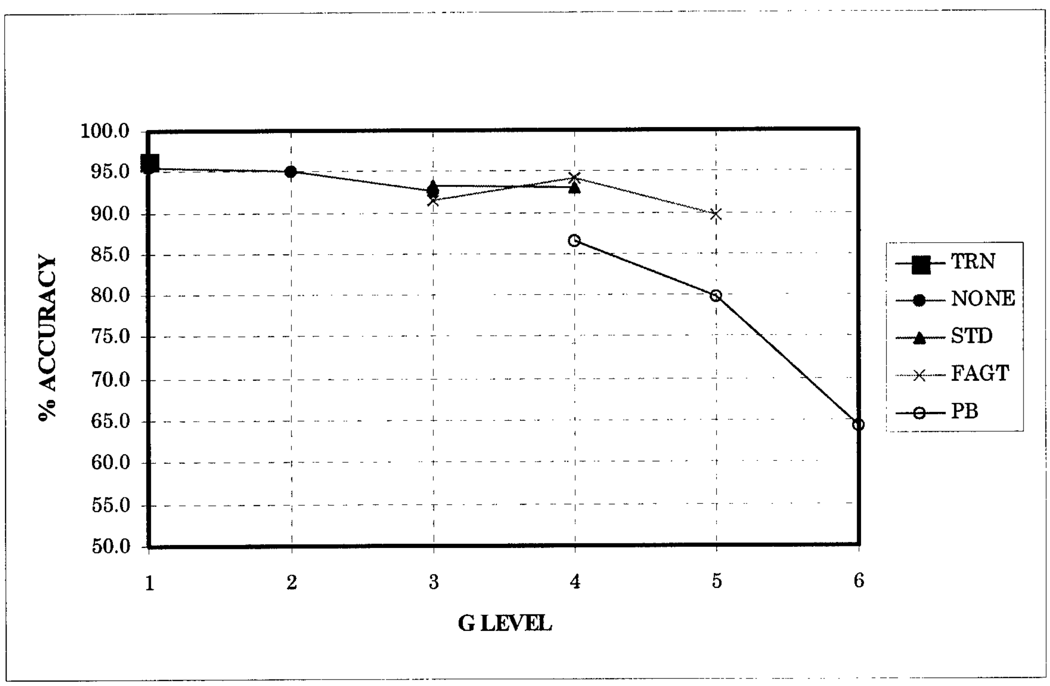


Figure 3. Average Word Accuracy vs. G

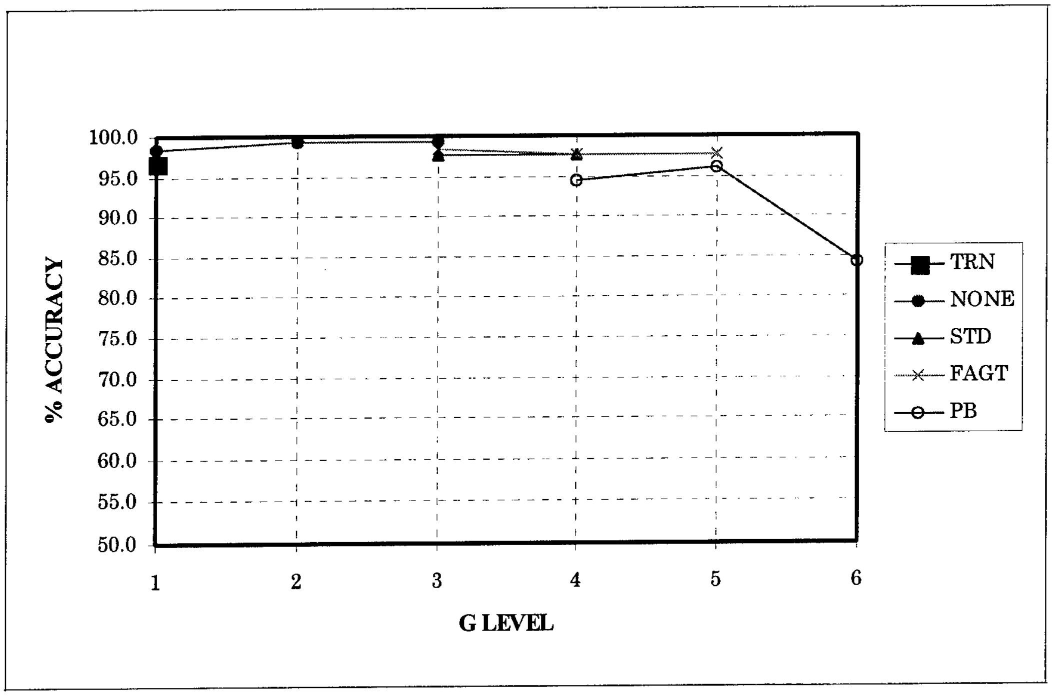


Figure 4. Word Accuracy vs. G, Speaker AP

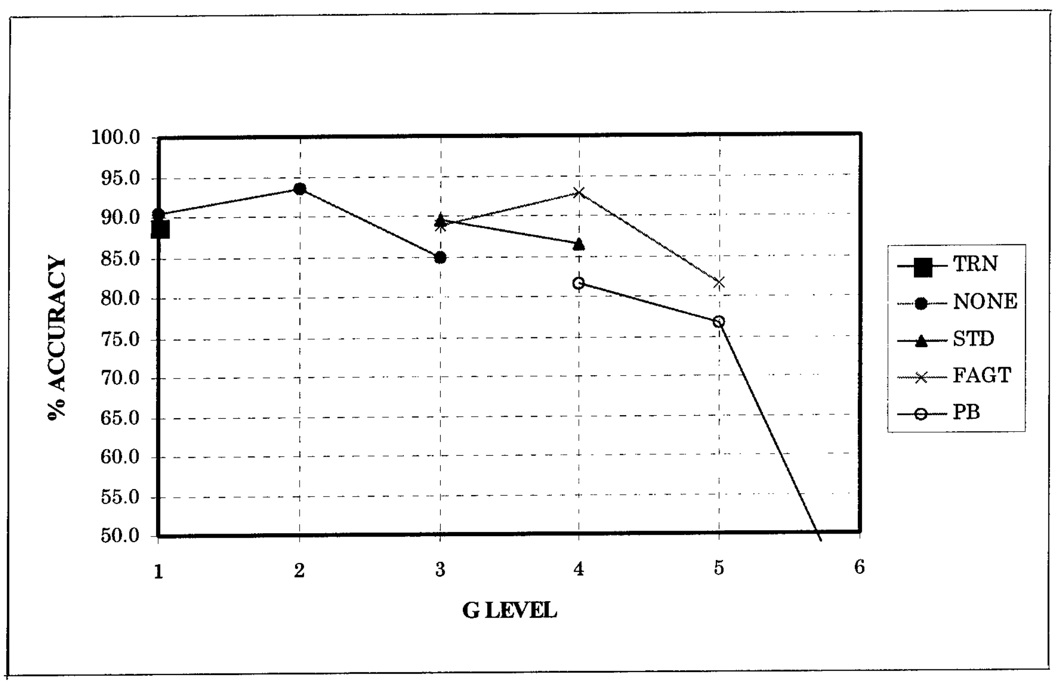


Figure 5. Word Accuracy vs. G, Speaker ML

FLIGHT TEST PERFORMANCE OPTIMIZATION OF ITT VRS-1290 SPEECH RECOGNITION SYSTEM

David T. Williamson
Timothy P. Barry
Kristen K. Liggett

Vehicle-Pilot Integration Branch
2210 Eighth St Suite 11
Wright-Patterson AFB, OH 45433-7521, USA

SUMMARY

This paper discusses the performance optimization of an ITT VRS-1290 continuous speech, speaker dependent speech recognition system that was flight tested in a NASA Lewis Research Center OV-10A aircraft. A 53-word vocabulary was tested with twelve pilots using an M-162 microphone headset on the ground and under 1g and 3g flight conditions. Digital Audio Tape (DAT) recordings were made of both the subjects' input and ambient background noise. Noise levels in the rear cockpit were in excess of 115 dB, with signal-to-noise ratios measured as low as 12 dB. During the early stages of the flight test, performance of the ITT system was poor, with some subjects achieving below 60% recognition accuracy. The DAT recordings became a critical element in the troubleshooting and optimization of the ITT system. A combination of input gain, noise calibration, and ITT recognizer engineering parameters were adjusted based on DAT testing to achieve an average word accuracy of 97.7% in the 1g condition and 97.1% in 3g across all subjects.

1 INTRODUCTION

Speech recognition has long been advocated as a natural and intuitive method by which humans could potentially communicate with complex systems. Recent work in the area of robust speech recognition in addition to advances in computational speed and signal processing techniques have resulted in significant increases in recognition accuracy, spawning a renewed interest in the application of this technology. Just recently, speech recognition systems have advanced to the point where 99% accuracy in a laboratory environment is commonplace. This high accuracy is key to acceptance of the technology by the user community.

The demands on the military pilot are extremely high because of the very dynamic environment within which they operate. Because workload is high and the ability to maintain situational awareness is imperative for mission success, voice control is ideal for military cockpit applications (ref. 1). Today's pilots must deal with vast amounts of information from both on-board and off-board sources. The single seat fighter pilot has limited ability to effectively manage all of the various information available using just hands and eyes. For these reasons, researchers have been exploring the possibilities of using speech recognition technology to augment the pilot's ability to control and display information in the cockpit (ref. 2, 3, 4, 5). In earlier research experiments, voice was used as a way to make discrete entries to cockpit systems, such as radios and navigation computers. But with newer, more reliable systems, the pilot can use voice control to naturally interact with the on-board systems in much the same manner as communicating with another crewmember. The use of voice command and response in the cockpit will provide an alternate means of controlling aircraft systems, receiving information from on-board computer systems, and managing off-board data. Other functions might include weapon manipulation, communications and navigation control. For instance, the pilot might ask the computer to display or remove certain information on the multifunction displays, change aircraft master modes, or get a verbal response to queries about fuel or weapons status. The idea of pushing numerous bezel buttons to access specific information on a multi-function display would be relieved by one verbal command, allowing the pilot to simultaneously keep hands on the stick and throttle and eyes out of the cockpit while using the voice system.

The potential use of automated speech recognition technology as a natural, alternative method for the management of aircraft subsystems has been studied by both the Air Force and Navy for over 10 years, but because recognition accuracies had not attained acceptable levels for use in the cockpit, this technology has not yet become operational in that environment (ref. 6, 7, 8, 9). There are a number of efforts that will contribute to the effective integration of a voice input and feedback system in the cockpit. One of the most important of these is the testing which will determine whether the technology is capable of operating in high noise, high g, and high vibration found in today's aircraft.

This paper discusses the environmental flight testing of an ITT VRS-1290 speech recognition system in an OV-10A aircraft. This system was chosen based on favorable results obtained in previous evaluations (ref. 10, 11, 12). In conducting the test, all speech was captured on high quality digital audio tape (DAT) to be used for subsequent testing of other speech recognition systems in the laboratory. As it turned out, these DAT recordings were critical in the performance optimization of the ITT system which lead to a successful outcome from this flight test effort.

2 OBJECTIVE

The objective of this experiment was to measure word recognition accuracy of the ITT VRS-1290 speech recognition system in an OV-10A test aircraft both on the ground and in 1g and 3g flight conditions. A secondary objective was the collection of a speech database that could be used to test other speech recognition systems.

3 TEST CONFIGURATION

Test Aircraft

The aircraft used for this experiment was an OV-10A aircraft operated by NASA Lewis Research Center in Cleveland, OH and is shown in Figure 1. This aircraft was a twin engine, two crew member, tandem seating turboprop aircraft. The OV-10A was capable of pulling up to 5.5 gs, but due to equipment constraints the test profiles were limited to 1 and 3g maneuvers.

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

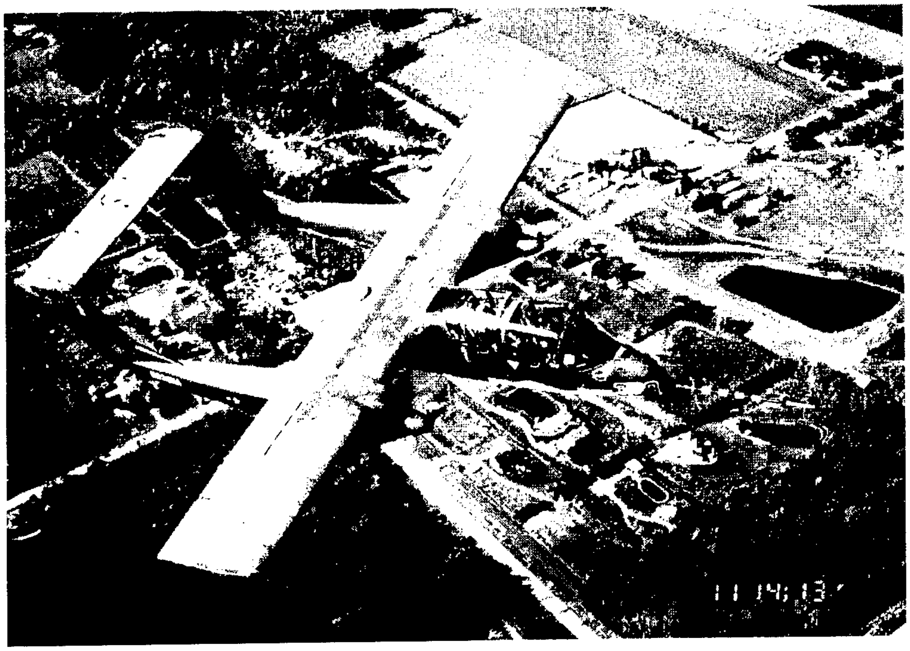


Figure 1. NASA LeRC OV-10A Test Aircraft

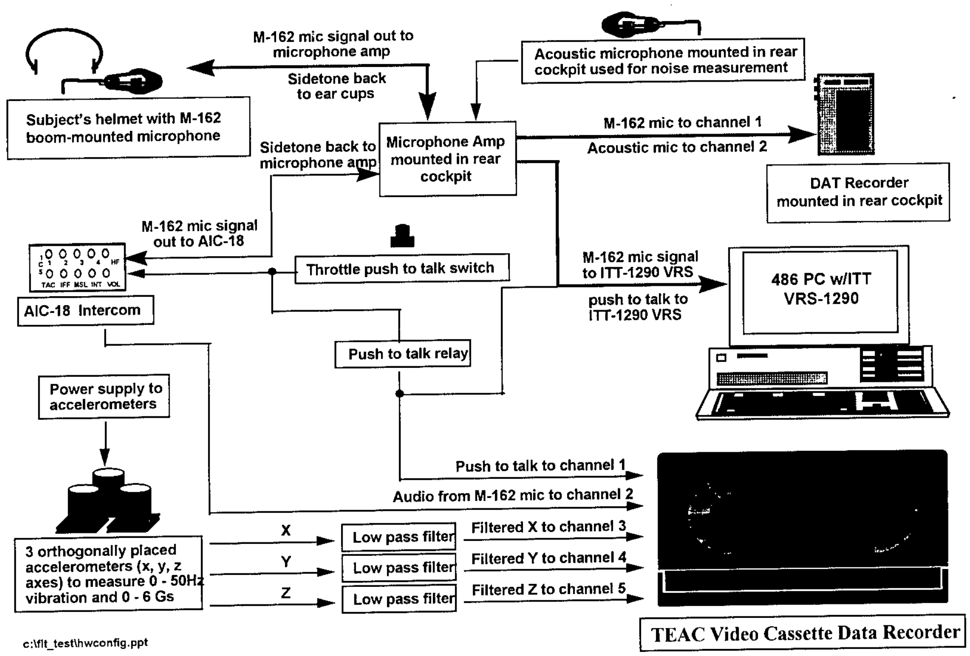


Figure 2. Flight Test Instrumentation

Flight Instrumentation & Test Hardware

An avionics test rack, located in the cargo bay of the aircraft, contained all the necessary research equipment. An IBM PC-compatible 80486 ISA bus computer system hosted the ITT VRS-1290 board and was mounted in the avionics rack along with a 24-track VHS data recorder. Figure 2 represents the instrumentation used during the flight test.

Audio Distribution and Recording

To provide the necessary audio to the speech recognition system, intercom, and recording equipment, a custom audio interface box was mounted in the rear cockpit. This box was designed and built under a joint effort between Armstrong and Wright Laboratories. The audio box was designed to take a variety of inputs from standard Air Force microphones including the M-162 boom-mounted microphone used in this test as well as M-87 headset and M-169 oxygen mask microphones. Input and output was also provided for

connecting a standard intercom system for sidetone feedback as well as communication between the subject and the NASA pilot in the front seat. The audio box also allowed the input of a Knowles BL 1994 acoustic microphone which was used to record ambient background noise inside the cockpit. A portable DAT recorder was mounted in the rear cockpit to obtain high quality recordings of both the subjects' speech and the ambient noise.

Speech Recognition System

The ITT VRS-1290/PC Voice Recognizer Synthesizer system was used for all speech recognition tasks. The VRS-1290 is a speaker-dependent device which uses a Template Determined End Point (TDEP) speech processing algorithm to provide continuous speech recognition of up to 500 unique words at any one time, with a total capacity of 2,000 words. Although not intended for an airborne environment, the system functioned well in the rack-mounted PC.

Software

The ITT TGS (Template Generation System) program supplied with the speech recognition hardware was used to "train" the subjects (for template generation). Custom software written in-house was used for prompting the subjects, performing the speech recognition and subsequently recording the recognition results on the PC.

Vocabulary/Grammar Structure

The vocabulary consisted of 53 words and phrases that represent various tasks that could be accomplished in a military aircraft. The vocabulary is shown in Table 1. The 53 vocabulary words and phrases were combined to form 91 test utterances to be used during ground and flight test conditions. Examples of the test utterances are presented in Table 2. The words in parentheses were intended to be optional words that would not be necessary to determine the meaning of a particular utterance. The words separated by "/" indicated synonymous words that would accomplish the same objective. This was designed into the test vocabulary to illustrate a more flexible interaction between the speech system and the subject.

North	**Show**	**Degrees**	**To**	**Before**
South	**Zero**	**Point**	**T-F**	**After**
East	**One**	**Minutes**	**T-A**	**I-D-S**
West	**Two**	**Ten**	**Ground-Map**	**Comm**
Range	**Three**	**Twenty**	**Pencil-Beam**	**Flt-Director**
Change	**Four**	**Forth**	**Weather**	**Radar**
Delete	**Five**	**Eighty**	**Beacon**	**Flight-Plan**
Modify	**Six**	**One-Sixty**	**North-Up**	**Page**
Add-new	**Seven**	**Two-Forty**	**Track-Up**	**Layer**
Goto	**Eight**	**Radar-mode**	**Heading-Up**	
Display	**Nine**	**N-R-P**	**Sector-Up**	

Table 1. Flight Test Vocabulary

Action Desired	Sample Utterance
Lat/Long Data Entry	**North 7 5 degrees 2 4 point 1 minutes**
Radar Range Setting	**Range twenty**
Map Orientation	**Change Display (to) Heading-Up**
Radar Mode Changes	**Change Radar-mode (to)/Give-me Beacon**
Display Selection	**Display/Show/Go-to Radar (page/layer)**
Add Nav Reference Point	**Add-new N-R-P after 3 4**

Table 2. Example Test Utterances

3 TEST PROCEDURES

Subjects

Sixteen subjects took part in this study. However, due to malfunctioning of the recording equipment on four subjects' flights, only twelve subjects had complete DAT audio data for all flight conditions. Eight of the twelve subjects were recruited from Wright-Patterson Air Force Base (WPAFB). All were rated military pilots. The remaining four subjects were NASA LeRC OV-10 pilots, each with prior military experience.

Experimental Design

The experimental design was a single-factor Within Subjects design with five levels of the Environment independent variable. All subjects were tested in the following conditions:

1) **Lab** - 182 utterances spoken in the laboratory environment
2) **Hangar** - 182 utterances spoken in the aircraft in the hangar with no engines running
3) **1g1** - 182 utterances spoken in the aircraft while flying wings level
4) **3g** - 67 utterances spoken in the aircraft while pulling 3 gs
5) **1g2** - 182 utterances spoken in a second 1g condition to test for possible fatigue effects.

Lab Testing

Participation was divided into two separate sessions. The first session consisted of generating the subjects' templates in a laboratory setting and collecting some baseline performance data. Subjects were briefed on the nature of the experiment and performed template enrollment. An identical system to the one in the aircraft was used as the ground support system for template generation. The subjects used the same helmet and boom-mounted microphone that was used in the aircraft. Template training involved the subjects' speaking a number of sample utterances which were prompted by the TGS software package delivered with the ITT hardware system. Once template generation was completed, a recognition test followed which consisted of reciting the utterances to collect baseline recognition data. Each of the 91 vocabulary utterances were spoken twice for a total of 182 utterances spoken in the laboratory. All of the laboratory training and testing utterances were recorded to allow subsequent testing on the ITT system or testing of a new speech recognition system.

Aircraft Testing

The second session began with a cockpit briefing that covered the operation of the test equipment (starting the DAT recorder, placement of the microphone, etc.), and safety issues. The subjects were provided with a knee board that contained the various checklists and a printout of the utterances to be spoken during the flight test. This printout was provided as a back-up in case of equipment problems.

During data collection, both on the ground and in the air, subjects sat in the rear seat of the OV-10A and were prompted with a number of utterances to speak. All prompts appeared on a 5" x 7" monochromatic liquid crystal display in the instrument panel directly in front of the subject. The recognition system attempted recognition after each spoken phrase with the recognition result stored for later analysis. DAT recordings were made of the entire data collection session.

The first aircraft test session was performed in the hangar to provide a baseline on the aircraft in quiet conditions. This consisted of each subject speaking the 91 test utterances twice -- for a total of 182 utterances. During both ground and airborne testing, subjects needed little or no assistance from the pilot of the aircraft. Close coordination was required, however, between the pilot and subject while the 3g maneuvers were being performed.

The flight test profile consisted of three conditions: (1) straight and level flight (1g), (2) 3g flight, and (3) repetition of the 1g condition to examine potential fatigue effects.

4 PRELIMINARY RESULTS AND PERFORMANCE OPTIMIZATION

During the first several flights, ITT word accuracy was well below expectations at around 55%. In the course of investigating potential causes for this poor performance, several problems were discovered. These problems were primarily audio related but also had to do with a lack of understanding of some of the engineering parameters that controlled the ITT system. Two such parameters were Noise Tracker and Noise Tracker Rejection flags that were both enabled. These parameters were primarily designed to enable rejection of spurious impulse noises, such as door slams. With the noise tracker parameters enabled, the system too often rejected valid utterances as noise, especially utterances at a terminal node of the grammar. Disabling these parameters resulted in at least a 10% improvement in word accuracy.

Another problem occurred when the subjects were required to perform a calibration of the system prior to a given flight condition. This process performed two functions simultaneously: background noise calibration and automatic gain control (AGC) parameter setting. During noise calibration, the system prompted the subject to be quiet for a short period. During this silence period, the system would generate a template of the background noise to adjust the voice templates for use in the higher noise environment. Due to procedural problems, however, this noise calibration was sometimes bypassed, resulting in poor inflight recognition performance. During AGC adjustment, the subjects were required to speak the phrase "ONE TWO THREE FOUR". After this digit phrase was spoken, the system would adjust the gain up or down and repeat the process until it was satisfied with the audio level. Most of the time, however, the system would freeze during this calibration step, requiring the subject to restart the computer. In order to ensure an accurate noise reading, the gain had to be fixed at a specific level. This required extensive retest with the DAT tapes generated in flight to find an optimum gain setting. It was discovered during the course of gain optimization that the flight system was being overdriven. Once the input audio gain to the ITT was reduced, performance again improved dramatically.

Figure 3 shows the results of this optimization process. LIVE1 is the result of three subjects' in-flight performance before optimization. DAT1 is the retest using the DAT recordings of the same three subjects with proper noise calibration and lower input gain. DAT2 is the final retest using the same three subjects' DAT recordings, this time with the Noise Tracker parameters turned off. The final data point, LIVE2, is the in-flight average word accuracy of five subjects that flew with the optimized configuration.

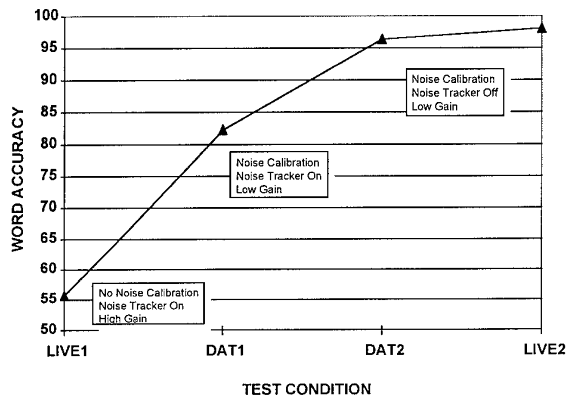


Figure 3. ITT Performance Optimization

5 FINAL RESULTS

Due to the audio and system problems encountered during the experiment, only five of the sixteen subjects had valid real-time recognition performance data in-flight. Four of the sixteen subjects experienced problems with the DAT recording equipment, resulting in unusable or non-existent audio data. Audio recordings were successfully collected for a total of twelve subjects in the study.

The data analyses were done in two stages. The first stage involved a comparison of "live", in-flight word recognition performance with word recognition performance obtained by playing the DAT recordings made in-flight into the ITT system back in the laboratory. The premise was that if no significant differences were found between live vs. DAT performance on the five subjects that flew with the optimum configuration, then the remaining subjects with complete DAT audio could be retested in the lab in the same way. Figure 4 shows the mean word recognition performance for both live and DAT recordings for the five subjects who had valid in-flight data.

An Analysis of Variance revealed no significant differences in word recognition performance when providing the ITT system with both live and digitally recorded audio signals.

With no performance differences found between live and DAT audio signals, all of the remaining analyses were done using DAT audio tape as the input to the VRS-1290. This provided

complete recognition data for twelve subjects. Figure 5 shows the mean word recognition performance obtained for each of the test conditions.

Three comparisons were of primary interest:

1. Ground (Lab + Hangar) versus air (1g1 + 3g + 1g2) performance
2. 1g (1g1 + 1g2) versus 3g performance
3. 1g1 versus 1g2 performance

Orthogonal comparisons were done to make each of these comparisons. No significant differences were found for any comparisons.

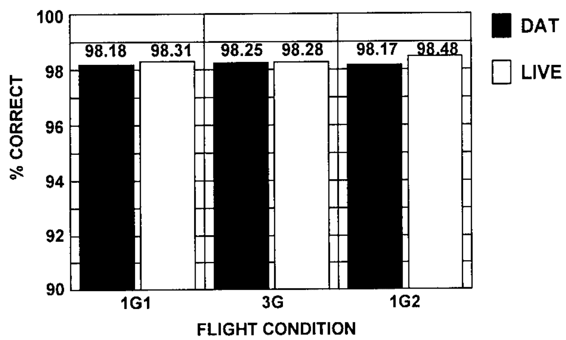


Figure 4. Mean word accuracy for live and DAT testing

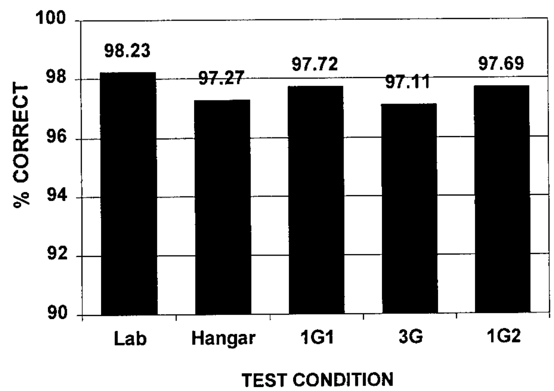


Figure 5. Mean word accuracy for each test condition.

6 DISCUSSION

Once the audio level and system parameters were optimized, the ITT VRS-1290 Voice Recognizer Synthesizer system performed very well, achieving over 97% accuracy over all flight conditions. It was anticipated that performance would significantly degrade in flight, especially in the 3g condition, due primarily to a decrease in the signal-to-noise ratio during this maneuver. Signal-to-noise (S/N) ratio measurements showed a 6 dB decrease in S/N in the 3g condition (18 dB S/N at 1g and 12 dB S/N at 3g. Once the background noise calibration was performed, the system was able to effectively compensate for the aircraft noise background and, therefore, no significant degradation was found.

This experiment highlighted several important lessons learned in the flight testing of automatic speech recognition systems. First, audio is everything. Microphones, pre-amps, intercom systems, automatic gain control (AGC), and aircraft wiring all must be carefully considered in order to achieve a successful speech interface. Air Force microphones vary greatly in quality, even among the same type. The M-162 microphone, however, turned out to be a good choice for this test. The custom pre-amp, which could easily be integrated as a small front-end circuit in an operational system, allowed the bypassing of the old AIC-18 intercom on the OV-10 which would have provided an unsuitable speech signal to the recognizer. Also, due to problems in the AIC-18, sidetone feedback to the subjects' earcups was virtually nonexistent. This made it difficult for the subjects to hear themselves speak to the recognizer. If the AGC circuitry on the ITT VRS-1290 performed as it should, the system would have adjusted the input gain to the appropriate level instead of having to be fixed at some predetermined level. Finally, the aircraft audio wiring to the speech system must be shielded and grounded properly to prevent 400 Hz interference on the speech signal.

The next lesson learned was that in order to get the maximum performance out of a speech recognition system, the application designer needs to be very familiar with the various parameters that control portions of the recognition process. This includes such parameters as rejection thresholds and, in the case of the ITT system, noise tracker rejection. By changing a Noise Tracker flag in the ITT from a 1 to a 0 word error rate was reduced by over 10%.

The final lesson learned was making sure the subjects were properly trained on the system. For speaker dependent systems, the vocabulary enrollment process is very important. If the templates generated from this process are not representative of how the subject will speak to the system in the aircraft, then performance will be degraded. Training a speech system is a two way process. The system learns how the speaker says a particular word or phase while at the same time the speaker learns how to talk to the system in a way that will maximize accuracy. The more experienced a speaker is with a particular system, the better the performance.

7 CONCLUSIONS

This flight test represented one of the most extensive in-flight evaluations of a speech recognition system ever performed. Over 5100 utterances comprised of over 25,000 words or phrases were spoken by the twelve subjects in flight. This combined with the two ground conditions resulted in a test of over 51,000 words and phrases. The audio database of DAT recordings will be transferred onto CD-ROM to facilitate laboratory testing of other speech recognition systems. The CD-ROM database will also be available for distribution to the speech recognition research community.

Another flight test is underway, this time on an OV-10D aircraft at NASA Lewis using a standard 12-P oxygen mask with an M-169 microphone. The speech database from this test will consist of sixteen subjects speaking a 47 word vocabulary under three flight conditions: 1G with a low power setting, 4Gs with 98% power, and a second 1G with 95% power. While the oxygen mask is providing at least a 15 dB attenuation of the background noise, the added breath noise and valve sounds from the oxygen mask are providing a challenge for this evaluation. Both an ITT VRS-1290 and Verbex Speech Commander system will be evaluated. Once again, all training, ground test, and flight test utterances along with the ambient cockpit background noise will

be recorded on DAT. This DAT database will then be transferred to CD-ROM for distribution to the research community.

The concept of speech recognition in the fighter cockpit is very promising. Any technology that enables a pilot to stay head-up and hands-on will greatly improve flight safety and situational awareness. It is hoped from this research that a robust speech recognition capability will emerge which can provide these operational benefits in the near future.

8 REFERENCES

1. Starr, A. F. (1993). Is control by voice the right answer for the avionics environment? In C. Baber & J. M. Noyes (Eds.), *Interactive Speech Technology* (pp.85-97). London: Taylor & Francis.

2. Lizza, G. and Goulet, C. (1986). Cockpit Natural Language. In *Proceedings of the 1986 National Aerospace and Electronics Conference* (pp. 818-819). Dayton, OH: IEEE.

3. Beckett, P. (1987). Voice Control of Cockpit Systems. In *Information Management and Decision Making in Advanced Airborne Weapon Systems: Conference Proceedings of the Aerospace Medical Panel Symposium* (pp. 27-1 - 27-8). Toronto, Canada.

4. Williamson, D. T. and Barry, T. P. (1990). Cockpit Application of Voice Technology. In *Proceedings of IEEE/AESS Symposium.* (pp. 57-60). Fairborn, OH: IEEE.

5. Cooke, N. (1990). RAE Bedford's Experience of Using Direct Voice Input (DVI) in the Cockpit. RAE-TM-FM-43.

6. Werkowitz, E. B. (1984). Speech recognition in the tactical environment: the AFTI/F-16 voice command flight test. In *Proceedings of Speech Tech '84 Voice Input/Output Applications Show and Conference* (pp. 103-105). New York, New York: Media Dimensions.

7. Williamson, D. T. and McDowell, J. W. (1986). The implementation of a voice actuated radio management system in a C-135 aircraft. In *Proceedings of AVIOS '86 Voice I/O Systems Applications Conference* (pp. 144-151). Alexandria, VA: American Voice Input/Output Society.

8. Williamson, D. T. (1987). Flight test results of the AFTI/F-16 voice interactive avionics program. In *Proceedings of AVIOS '87 Voice I/O Systems Applications Conference* (pp. 335-345), Alexandria, VA: American Voice Input/Output Society.

9. Loikith, G. and Hall, B. (1986). TF/A-18A Airplane Voice Interactive System Flight Test Demonstration. Navy TR SY-36R-86.

10. Sweider, R. F. (1987). Operational evaluation of voice command/response in an Army helicopter. In *Proceedings of Military Speech Tech '87* (pp.143-146). Arlington, VA: Media Dimensions.

11. Holden, J. M., Stewart, S. J., and Vensko, G. (1987). Speech recognition in a helicopter environment. In *Proceedings of Military Speech Tech '87* (pp.147-152). Arlington, VA: Media Dimensions.

12. Barry, T. P., Solz, T. J., Reising, J. M., and Williamson, D. T. (1994). The use of word, phrase, and intent accuracy as measures of connected speech recognition performance. In *Proceedings of the Human Factors Society 38th Annual Meeting* (pp. 325-329). Nashville, TN: Human Factors Society.

Hidden Usability Issues in Speech Interfaces

Malcolm J. Cook, Charles Cranmer, Carol-Ann Milton
Division of Psychology, School of Science
University of Abertay Dundee, Bell Street
Dundee DD1 1HG, UK

email: mcook@river.tay.ac.uk or
c.cranmer@river.tay.ac.uk

Robert Finan, Andy Sapeluk
Division of Power and Electrical Engineering
University of Abertay Dundee, Bell Street
Dundee DD1 1HG, UK

email: rfinan@river.tay.ac.uk or
a.sapeluk@river.tay.ac.uk

Telephone: + 44 1382 308749
Facsimile: + 44 1382 308877

Abstract

The increasing power and sophistication of speech recognition and speech synthesis has encouraged speculation that human factors problems in implementing speech interfaces will diminish dramatically as technological develops (Zue, Glass, Goddeau, Goodine, Hirschmann, Leung, Phillips, Polifroni, and Seneff, 1991). Advanced speech interfaces have been investigated and are currently being integrated into prototypes for advanced civil and military cockpits (Gerlach and Onken, 1993, Onken, 1995, Turner, 1995; Turner, 1996; Steeneken and Gagnon, 1996). The reason for the introduction of speech-based interfaces is to increase the time available for head-up flight and, thereby, to improve flight performance and safety. The advantage which is claimed for delivering information via speech-based interfaces is a reduction the vast quantities of information normally presented in visual displays in the cockpit and the release of the pilot from head down management of cockpit systems. Directly or indirectly, the benefits of splitting information delivery and data command/entry across modalities are often justified in terms of independent information processing. The independent nature of the processing in turn assumes there will be no interference between tasks or degradation in performance. Pilot research by the authors indicates that there are problems related to memory and workload that are present in current technology and will remain in future solutions with speech-based interfaces (Finan, Cook and Sapeluk, 1996). These problems will remain in even though recognition accuracy is increased because they reside in the limits of the human operator to manage multi-modal environments. In a simulated multi-task environment self-reports and performance at moderate to high levels of workload with multi-modal interfaces have shown that overall performance with speech-based interfaces is degraded. The use of multi-modal interfaces resulted in degraded performance on tasks requiring extended processing of information and recall of information from memory.

1. Introduction

It has been argued that speech technology can provide significant advantages in the design of aerospace cockpits if their use is restricted to low and moderate levels of workload (Cresswell-Starr, 1993). A major benefit of adopting speech interfaces is that they would significantly increase the amount of head-up time in the cockpit because instructions could be given to on-board systems and information received from the same systems without having to look down. The extra head-up time would, in turn, increase pilot performance and safety.

Those making these suggestions accept that the current rates of error and future performance make these systems unsuitable for real-time control. The same author (Cresswell-Starr, 1993) has accepted that users may need visual references to ensure the pilot are aware of current system status because they apparently forget the current mode. This problem in recalling the mode of operation cannot be remedied by adding information to an already cluttered head-up display system. Thus, in solving the problem of providing information to the pilot without requiring head-down operation one may inadvertently introduce another problem in terms of a extra burden on memory. This problem

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

could be equally problematic in other hands-free voice activated applications.

The arguments in favour of speech can be listed as follows:-

- Speech as an automatic over-learnt skill.
- Significantly greater head-free and hands-free operation.
- Use of additional unused processing capacity with no cost.
- Separability of information and information sources.
- Increasing power and sophistication of speech recognition technology.
- Restricted use in low and moderate levels of workload for well-defined tasks.

There are, however, a many problems and unanswered questions about the application of speech interfaces listed below:-

- The overall task may require integration of disparate information which is impeded by multi-modal presentation.
- Speech presentation may impose a greater memory burden as information presentation is transient.
- Switching attention between modalities may be slow and have a high cost.
- Auditory presentation may pre-empt and disrupt visual presentation.
- Recognisers fail as user's become stressed.
- Restricted vocabulary is unnatural and may have a high cost.

Anecdotal evidence suggests that using speech in multi-task environments can be problematic. Non-standard radio communication, which is composed of information which is poorly integrated into current tasks, has been cited as a contributory cause in aircraft accidents (Cheung, Money and Sarkar, 1996) and use of cellular phones while driving has been accepted as a possible cause of automobile accidents (Adams, Tenney and Pew, 1991). Thus, there are indications that the disruptive and pre-emptive effects of auditory presentation exist. Or, it may be that auditory and visual tasks carried out concurrently give rise to greater performance decrements when compared with individual task performance because tasks in different modalities use a common resource, like attention or working memory.

Implicitly it has been assumed that delivery of information via a speech-based interface would increase the bandwidth of information presentable but the evidence given above casts doubt upon this justification for speech-based interfaces. It is also assumed that additional information would accrue minimal extra costs in terms of information processing because of the separability of the information presented in terms of human information processing. This assumption of additional processing capacity without costs in performance is equally doubtful.

These assumptions with regard to human information processing are largely derived from the models of multi-task performance, experimental data and reviews of multi-task performance put forward by Wickens and colleagues (Wickens, Sandry and Vidiluch, 1983; Wickens, Vidiluch and Sandry-Garza, 1984; Wickens, 1984; Wickens and Flach, 1988; Wickens, 1989; Wickens, 1992; Stokes, Wickens, and Kite, 1990). Wickens's original model (Wickens, 1984) divided information processing into three stages: encoding, central

processing and responding. Information could be coded as spatial or verbal and presented to the auditory or visual modalities. The output or responses in the model could be manual or vocal. Maximum performance in Wickens' model of processing resources would occur if the code, modality or information presentation and the response were compatible. In the original model, there were two optimal routes. The first was auditory presentation with verbal encoding and verbal response, and the second was visual presentation with spatial encoding and manual response.

The experiments reported in this experiment are specifically directed at Wickens' model. According to the model speech-based systems could be introduced with little cost because the input and output were related via an internal verbal code. Again according to the model there should be little task-interference because these other tasks are delivered with visual presentation, encoded in spatial terms and requiring a manual response.

The quantification of the effects of workload in relation to simultaneous visual and auditory tasks like those carried out in the cockpit are largely unknown and Cresswell-Star's (1993) assertion that speech could be used at moderate levels of workload is largely unsupported by appropriate experimental evidence. It is very important to recognise the two components of the argument in support of multi-modal input/output: independent information processing and stimulus-response compatibility. Together these premises suggest that presenting information through different channels results in little or no extra cost in terms of quality of processing.

There are a number of flaws that have been identified in the arguments used to support speech interfaces. Firstly, the delivery of information across different modalities can have higher attentional costs than delivery utilising a single modality. This is because it takes greater amounts of time to switch between modalities than it does to switch attention within modalities (Wickens, 1989). This problem may mean that redundant coding in two modalities can improve performance in concurrent tasks but give higher error rates when events occur in different modalities in close temporal proximity (Cook and Elder, 1996). The same attention switching argument can equally provide a plausible explanation for aerospace accidents in which non-standard radio communication was identified as a contributory cause (Cheung, Money and Sarkar, 1996) and accidents related to the use of cellular phones in cars (Adams, Tenney and Pew, 1991).

Secondly, there is a tendency for auditory signals to pre-empt or disrupt simultaneous processing of visual information which may relate to the attention is directed by sound or differences in the length of the anatomical pathways in audition and vision. Whatever the reason, this pre-emption or disruption undermines the utility of additional information presented via the auditory channel. Interruption of processing is very significant and it has been accepted as a possible contributory cause in accidents (Adams, Tenney and Pew, 1991; Satchell, 1993; Woods, Johannessen, Cook, and Sarter, 1994). Recent research using carefully controlled experiments suggests that there may indeed be significant asymmetries in the attentional control of one modality over another, at least with respect to spatial information processing.

Thirdly, the segregation of the information processing may be advantageous for processing of specific items of information but it could interfere with the integration of information across modalities. The integration of information may be essential feature of higher-level decision making typical found in the tasks where volume of information indicates that speech-based interfaces could useful.

Fourthly, there may be optimistic estimates of the ability to use speech-related information because evaluations may have failed to reveal the suffix effect in auditory presentation. The suffix effect in psychological literature more frequently refers to the loss of the most recently presented information from acoustic short-term memory (Baddeley, 1994) when a delay occurs between the final item presented and recall. In the event-driven real-time tasks, which may attract developers of speech-based systems, tests of recall could artefactually be inflated by

the absence of the suffix effect and suggest that memory performance was more effective than it might be. It has been shown that the suffix effect can be caused interposition of visual or auditory items prior to recall. Thus, multi-modal information presentation will not be immune to the suffix effect.

In the past attempts have been made to compare speech to its nearest alternative, keyed input and with certain modifications of the vocabulary favourable results can be achieved with restricted vocabulary (Damper and Wood, 1995; Damper, Tranchant, and Lewis, 1996). However, it can be argued that these changes remove some of the value of speech technology as it is no longer natural language and the use of restricted vocabularies adds another cognitive burden a very busy operator. Even though practice can result in a high degree of automaticity one might argue that adds significantly to the cost of introducing the technology. For example, evidence on the analysis of communication using military terms suggests that speech is more verbose indicating the human user's need to embellish communication to improve emphasis and add additional information which is not directly available from the restricted vocabulary (Achille, Schulze and Schmidt-Nielsen, 1995). Producing information rich communication may be both for the benefit of the listener and as a method of self-review and monitoring one's own knowledge. In simple terms, a restricted vocabulary does not have the qualities of a natural language used as a native tongue since birth, which is an inherently slow but effective method of communication. What may be required is a re-design of the task and not the addition of a problematic speech-based interface.

Many authors have questioned the use of speech input in non-critical and non real-time applications, and it has been suggested that redundant coding is necessary to prevent mistakes (Rudnicky and Hauptmann, 1992; Cresswell-Starr, 1993). If redundant coding is required this casts doubt upon the value of speech as a separate channel of information. This is all the more pertinent, when one considers that more moderate advocates of speech-based interfaces in the cockpit recognise difficulties with short-term memory recall and identify the need for visual feedback while operating with speech-based interfaces (Cresswell-Starr, 1993). The need for confirmation and the problems with recalling mode of operation indicate attention or memory resource problems or both. In personal discussions with regard to the MOD reports, which Cresswell-Starr (1993) used, it became clear that operators would frequently write information onto a knee-pad prior to data entry to prevent mis-remembering. What is particularly disturbing in this information is the use of pilots in the validation work because this suggests that the target population for which the systems are intended will use alternative strategies, to overcome the memory burden, which may create other performance decrements.

Thus, three pieces of evidence seem to indicate that memory load may be a hidden problem in using speech-based interfaces. First, Cresswell-Starr (1993) noted that operators would often forget the mode the system was currently in while using speech-based interfaces. Second, operators would often use a knee-pad to note down figures from speech systems prior to input. Third, a previous experiment by the authors (Finan, Cook, and Sapeluk, 1996) indicated that subjects forgot significant flight parameters with a frequency that could cause problems in terms of situational awareness (Cook, Cranmer, Finan, Sapeluk, and Milton, 1996). Altitude violations caused by situational awareness problems have been recently cited as a cause of a number of in-flight accidents and incidents (Raymond and Isaac, 1995) and they were one of the operator errors which the technology demonstrator CASSY (Onken, 1995) was originally meant to prevent.

The observations, listed above, suggest that pilots may be faced with a difficult choice of trying to maintain information in short-term memory or losing head-up time because of the need to write notes or confirm information received from auditory channels. In operational terms they may forget important information or fail to process important signals. Interestingly, these observations are not in accord with reports from test-pilots, in current projects (Turner, 1995; Turner 1996), which indicate the benefits of speech-based interfaces for

certain tasks. Any difference in the reports of pilots using speech-based interfaces and those found in experiment investigations may reflect the differences in the tasks selected for voice control, the relative simplicity of the information in input and output, the dialogue structure of those tasks or the remediating effects of practice. However, there often exist significant discrepancies between opinions with regard to performance and actual measured performance. The ability to self-monitor may be much reduced in the moderate and high demand tasks and errors in those situations are often attributed to poor decision making based on inadequate or incomplete knowledge of the current situation. In the aerospace community knowledge of the current system status is termed situational awareness. thus, it is hard to see how those making the errors can be aware of potential errors or omissions in their current knowledge.

The following series of experiments examined the ability of subjects to use speech-based interfaces for a limited range of tasks while controlling a notional four-engined aircraft. The significance attached to workload in Cresswell-Starr's (1993) report meant that three levels of workload in visual vigilance tasks, in conjunction with speech-based interface interactions. Visual vigilance tasks were used to mimic the sustained demand in cockpit tasks.

The primary aim of the first four conditions reported was to establish if the presence of a concurrent speech task had a deleterious effect on performance of a concurrent visual task. Predictions following Wickens' model (1984) would assume little difference in concurrent performance of visual and auditory tasks which are matched for stimulus presentation, encoding and response compatibility.

Thus, the key element of this research is the manipulation of demand on central resources by changes in the nature of the visual discrimination task and changing demands on memory. This type of manipulation allows the experiments to address Cresswell-Starr's (1993) suggestion that speech interfaces would be used in a restricted, well defined and stable set of tasks. A significant fall in the visual task carried out concurrently with speech-based tasks would indicate a conflicting demand on a common memory or attentional resource. This series of experiments is not designed to differentiate between these possibilities but it should indicate if interference occurs. Interference would indicate a major usability issue that is likely to remain after the development of more advanced recognition software and greater robustness of recognition performance.

2. Method

2.1 Subjects

Ten subjects were used in most of the experiments. A core seven subjects completed all of the experiments and other subjects were drafted in to increase the sample size. All the subjects were given preliminary training to familiarise them with the task prior to the start of the experiment.

2.2 Equipment and Software

Dialogue Designer is a Windows application (Finan, 1994a, 1994b), developed for A.T. and T. (GIS) Dundee, for design and testing of speech based human/machine interface dialogues. Dialogue Designer is a rapid-prototyping system for the development of speech based interfaces. It was designed to examine interaction protocols, error correction and task structure, which have been identified as important features of dialogues (Rudnicky and Hauptmann, 1992).

The visual vigilance tasks used in this experiment had previously been used to collect data for experiments on visual vigilance performance and multi-modal performance (Cook and Elder, 1996). The previous research provided a baseline set of data for the task variants and it was felt that this would prove useful in establishing the effect of introducing a concurrent speech task. The software for the visual vigilance tasks and the Dialogue Designer prototyping software were run on a 486DX2 Western Systems PC. The voice recognition software used was

the Voice Assist package supplied with Creative Labs Inc. Soundblaster 16 Bit SCSI-2 hardware.

2.3 Procedure

Three levels of task demand could be produced by changing one of the two visual vigilance tasks the subject was required to complete. Each subject was required to complete two conditions for each level of task demand. In one of the conditions the speech task was required concurrently and in the control condition it was not. Thus, for core subjects completing the basic series of experiments there were six separate conditions, speech present or absent with one of three levels of visual task demand. These conditions are shown in Table 1.

Both the visual tasks were displayed on the same PC monitor at a comfortable viewing distance in front of the subjects. The visual tasks required a speeded discriminative response to on-going visual events which appeared on the screen for 600 ms. Each presentation was followed by a blank period of 400 ms before the appearance of the next presentation.

The first task required the subjects to monitor the orientation of an arrow and the subject would indicate when the arrow pointed downwards by a keypress. This event occurred infrequently on ten percent of the presentations. The monitoring of arrow orientation is, hereafter, referred to as the *arrow task*.

The second visual task required the subject to monitor a display of blue and red squares that overlapped the arrow and to press a key when a significant event occurred. The monitoring of the pattern of squares in the low, moderate and high demand conditions is, hereafter, referred to as the *squares task*. In the low demand condition, the pattern of red and blue squares was repeatedly presented and a significant event was signalled by all of the squares appearing as red. In the moderate demand condition, the pattern remained constant for long periods of time and a significant event occurred when a single square moved position for a single presentation. In the high demand condition, the subjects watched a constantly changing pattern of red and blue squares and a significant event occurred when the pattern remained constant across two consecutive trials. The pattern of squares on significant events and regular presentations is illustrated in Figure 1, 2 and 3. In all cases the subjects were required to make a response within a second of pattern onset during a significant event in the arrows and squares tasks, and late responses were rejected.

To summarise, at each level of demand the subjects were required to carry out three tasks simultaneously or two tasks simultaneously. Either they would carry out the two visual vigilance tasks on their own or in conjunction with the speech task.

The number of correct detections, false alarms, and misses for the critical events were calculated for both visual vigilance tasks by manual assessment of a computer log generated by the experimental software. The possibility of order, fatigue and time of day effects in the vigilance tasks was accounted for by using the control run with only the visual vigilance tasks. The control runs were taken before or after the experimental session in a counterbalanced manner. The control data were used as a baseline measure of performance.

The third task that subjects performed in the experimental sessions required them to interact with a speech interface. The same computer provided prompts from a notional ground controller and handled speech input. Using a number of structured dialogues, some which are illustrated in Figures 4, 5, and 6, subjects could modify up to five types of system parameter. These parameters were altitude, bearing, radio frequency, engine and undercarriage status. Subjects were instructed to view these dialogues as the primary task, with the visual vigilance tasks being of secondary performance.

An important feature of the dialogues was the requirement to repeat the key input parameters in the input dialogue. This repetition was included to help subjects in recalling the relevant status parameters. Subjects were only tested on parameter

settings at the start of the experimental run and the end of the experimental program in the basic experiment. In a repeat of the high demand condition subjects were given a memory probe task to test situational awareness during the experimental run.

Subjects completed a detailed questionnaire at the end of each session with and without the speech task. The questionnaires required subjects to indicate the level of task difficulty for all three tasks, and an assessment of their competence with regard to each task.

In an additional, fourth condition the third condition of the experiment was repeated and during the second set of trials a memory probe task was used to establish if the subjects could recall the current flight parameters during the experimental run. The probe information consisted of a simulated report of another aircraft flying on a particular heading and bearing, tuned to a particular frequency. The subject's task was to decide if any of the parameters matched those their own aircraft. If the current heading and bearing of their aircraft and that of the other aircraft matched they requested a collision avoidance speech dialogue. Similarly if the radio frequency matched they could request a change in that parameter.

2.4 Experimental design and analysis

The small sample size, the types of data collected and the use different subjects across different trials, meant that a multivariate model would be inappropriate. Therefore, the data were analyzed with a non-parametric test, the Wilcoxon Matched Pairs Signed Ranks Test supplied as part of Minitab Version 10.1.

3. Results

3.1 Performance on Visual Vigilance Tasks

As Table 2 indicates the mean number of errors on the arrow task increases as demand increases when speech is absent but the pattern of errors with speech present is paradoxically reversed. The pattern of errors is, however, inconsistent across subjects. The absence of consistent trends in the arrow task is in direct contrast to the pattern found in the performance of the squares task. In the squares task the number of misses increased as the level of task difficulty and demand increased. There was an interaction apparent in the data across demand levels between speech present and speech absent conditions and this interaction was consistent across most of the subjects. The interaction was apparent as a much steeper increase in error rates, as demand increased, in the speech present compared to the speech absent condition. This interaction is shown Graph 1.

There were no significant differences in the pattern of false alarms either visual task in any condition. See Table 3 for the data on misses.

Subjects were relatively poor at judging their own performance on the arrows and squares tasks. This is highlighted in Table 4, which provides a comparison of subjects' own estimates of their performance and mean percentage misses. There was a trend for subjects assessment of their performance to fall between speech absent and present conditions. This perceived fall in performance was in line with the actual trends for the squares task.
However, there were dissociations in the perceived and actual performance in the arrows task.

3.2 Speech Recogniser Performance

The performance of the speech recogniser fell in the highest demand condition. However, individual performance on the concurrent tasks appears to be largely unrelated to the number of errors of the speech recogniser with speech for individual subjects.

Analysis of individual subjects performance showed that for subject 5, in the medium and high demand conditions, had very few problems in terms of speech recognition errors but the number of misses on the squares task rose dramatically. By contrast, subject 6 had a very high number of speech

recognition errors in both conditions but still experienced a dramatic rise in mean percentage misses in the squares task. It is equally important to note that most of the speech recognition errors were attributable to a small number of subjects.

As Table 5 indicates, the greatest increase in recognition errors was in terms of failed recognitions. The number of mis-recognised items was relatively constant across trials.

3.3 Performance on Speech-Based Tasks

Subjects rating of the difficulty of speech dialogues was consistent across the three conditions, even though there were more speech recognition errors in the high workload condition. Combination dialogues, where the notional ground controller required the subjects to change two parameters instead of one, were rated as most difficult in all three conditions. The performance ratings are shown in Table 6.

Subjects mean recall of initial flight parameters was consistently better for altitude than bearing and frequency. Frequency was consistently the most poorly recalled initial flight parameter for the three conditions. Figure 7 gives the recall performance for initial flight parameters and figure 8 the recall performance for the final flight parameters. The final altitude, bearing and frequency were consistently recalled at much lower levels in all conditions when compared to initial flight parameters. However, recall performance did increase between the low and moderate demand conditions for all parameters. Altitude was the best recalled of the final parameters.

Recall of the engine status and the undercarriage status which were binary descriptions of on/off and up/down was always relatively high with an indication of a ceiling effect in performance. These results are shown in figure 9.

When recalling flight parameters the subjects' incorrect responses could be coded in three ways. Either, no answer was given, those where an answer was correct in all but one digit, and those where answers given were incorrect in two or more digits. The results of this coding are shown in Table 7. Recall with no answer given or one digit incorrect were common in the low demand conditions. Recall with two or more digits incorrect were common in the high demand condition. By combining the failure to answer with the recall of two or more incorrect digits the high demand condition was found to contain the greatest level of recall errors.

3.4 Performance In Probed Recall Task

In the probe task, there no significant changes in the performance of the arrow task with speech absent or present. There were significant differences in the performance of the squares with speech absent or present and this was in terms of the number of misses ($n=8, t=0.0, p<0.05$). As shown in table 8, the absolute levels of performance in both visual tasks with and without probed recall were not significantly different.

The pattern of false alarm rates for both the arrows and the squares visual tasks in the high demand condition with and without probed recall were also similar. Table 9 shows the mean percentage of false alarms for the visual tasks in the two conditions.

The accuracy of subject's assessment of their own performance on the arrows task was poor, as shown in table 10. For example, for the high demand condition without the probe, performance in the arrows task with speech was rated as very poor (median=1) and the mean percentage misses was 5.6%. However, in the high demand with probe condition performance was similar 5.5% and the same task was rated as average (median=3). Subjects appeared to be more accurate at rating their performance in the squares task, in that they rated the no speech condition as poor (median=2) and the speech condition as very poor (median=1).

As shown in table 11, the pattern of speech errors for the probe recall task were generally very similar to those without the memory probe.

Subjects rated the speech dialogues as being slightly more difficult for altitude and frequency in the high demand with probe condition, than the high demand without probe. However, overall difficulty rating across the two conditions were generally similar. Table 12 shows the median scores for ratings of dialogue difficulty for the two conditions.

The recall of initial altitude and frequency were better in the high demand condition without probe (see figure 10) and recall of initial bearing was better in the high demand condition with probed recall. Final parameter recall exhibited a similar mixed pattern of results with final altitude recalled best in the probe absent version of the high demand task and final frequency was best recalled in the memory probed version of the high demand task. The final bearing was recalled equally well in the tests with probed recall absent and present (see figure 11).

Table 13 shows additional final flight parameter recall results which were effectively 100 percent correct in both conditions.

There were very few errors in the probed recall task. Figure 12 shows the number of incorrect responses given for the 1st set of probe dialogues. Figure 13 shows the number of incorrect responses given for the second set of probe dialogues. In absolute terms, subjects made ten incorrect responses from a possible total of sixty. More errors occurred towards the end of the experimental session for both sets of dialogues.

4. Discussion of Results

4.1 Interpretation of data in multi-task experiments

The performance in multi-task experimental paradigms is very difficult to control because subjects can select different strategies for dealing with task demands in ways that will affect recorded performance. Therefore, it is necessary to examine performance across all the full range of tasks attempted to try and identify the shifting emphasis in resource allocation across tasks. In this experiment this meant examination of performance on the squares tasks and the arrows task when they were and were not required to carry out in conjunction with speech tasks.

This detailed examination of performance is clearly the only way to identify the shifting emphasis in resource allocation across tasks because subjects' assessments frequently did not correlate with actual performance on the task assessed. The general pattern of subjective estimates of performance seemed to be more effectively driven by the memory demands of the task they were required to do. Thus, subjective performance assessments seemed to be mainly driven by the squares task in the first three conditions but in the probed recall version of the high demand task the need to retain information appeared to affect subjective estimates of performance. The subjective estimates for the difficulty of dual dialogues, in which subjects were required to maintain information for a period of time between two successive input dialogues, are in accord with this view.

4.2 Degradation of performance in high demand conditions

It appears that subjects experienced more difficulty with the squares task when accompanied by a speech task. Although the differences were not significant in the low and moderate demand conditions, subjects' error rates were significantly greater when speech dialogues were required in the high demand condition. There are two possible explanations for this. Firstly, increased competition for central resources in terms of working memory. Secondly, the effect of attention grabbing and interruption by the auditory input (Wickens, 1989).

The presence of greater error rates in simultaneous tasks is significant, because models of multi-task performance would suggest that performance should not degrade if information processing channels are distinct and compatible. The channels used by visual tasks and speech-based tasks in the current experiments seem to be in accord with those general principles. As far as possible the presentation modality, the preferred internal coding and the response mode for each task are distinct

and compatible.

The task interference at high levels of demand indicates that speech-based interfaces must be carefully tested to ensure that performance, on visuo-manual tasks, in the integrated multi-task multi-modal situation are not degraded by such interference effects. A key feature of these interference effects is the increasing impact with increasing demands on memory and sustained attentional resources. The measurement of task demands or workload is problematic, however, as there is no objective way of measuring the demand on the pilot in real-time to prevent the degradation of performance as workload rises. Indeed, continuous pilot workload monitoring has been a subject of general interest to the aerospace community and there are no definitive answers as yet (Satchell, 1993).

An important feature of the tasks in this study is the separability of the tasks and it is possible that interference effects will be greater when information must be integrated. However, the anecdotal evidence suggests that accidents develop when signals are presented in different modalities and are carrying information with regard to unrelated concurrent tasks (Adams, Tenney, and Pew, 1991; Cheung, Money and Sarkar, 1996).

It is interesting to note that the subjects missed as many events on the arrow task with/without speech in this series of experiments as the subjects in Cook and Elder (1996). It is also interesting to note that performance on the high demand tasks was significantly worse when compared to that recorded in Cook and Elder (1996). The dramatic increases in error rates for the squares task accompanied by speech seem to support the contention that the additional speech tasks contribute to degraded visual vigilance performance.

There remains the possibility that subjects are simply shedding the squares task as the most demanding of the three tasks they are attempting simultaneously. The fall in false alarm rates in the moderate and high demand squares task data certainly are in accord with this hypothesis but they are not significantly different in the speech present and speech absent conditions. The slight drop in false alarm rates could be interpreted as symptom of a depressed response because of inattention to the squares task. One might expect an improvement in the arrow task performance that would benefit from the released resources if that were the case but the results from the arrow task at high demand levels show no benefit in performance. Thus, it seems that much of the capacity that might be taken away from the squares task at high demand must be diverted to the speech tasks. This interpretation suggests that the additional speech task is placing a heavy demand on the subjects and the benefits of task shedding are lost to the resource hungry speech dialogues.

Although probed recall did not affect the general performance of the arrow and squares task, the subjects reported greater difficulty in performing that variant of the high demand task. Recall performance of the flight parameters was significantly improved by the subjects need to maintain an accurate model of the aircraft status in order to take appropriate action in the collision avoidance dialogues. However, it remains to be seen that this could be supported for higher event rates in the visual tasks or for prolonged testing. Problems in maintaining an adequate level of performance over prolonged periods or at high event rates clearly have implications for the application domain.

4.3 Arousal and Performance

Paradoxically, the initial series of experiments indicates that subjects appear to be better at recalling flight parameters in moderate and high demand conditions. However, subjects seem to make more catastrophic recall errors in high demand conditions for final flight parameters. In fact, many of the errors in all conditions were in the reporting of the final flight parameters.

The short delay at the end of the experiment prior to reporting could have contributed to the decrement in recall for final flight parameters but it seems unlikely that this is a suffix effect. The

subjects could have had the final flight parameters available in working memory for up to a maximum of two minutes and a minimum of 30 seconds before the end of the experiment. The time available should allow ample time for encoding and storage in a more permanent form. There are at least three problems that may have prevented perfect recall. Failure to recall could reflect poor storage, poor retrieval or inadequate marking of the temporal order of data. The last presents an intriguing possibility that users would mis-remember previous information as the most recent information which could result in faulty decision making. It is likely that all three mechanisms play a role in poorer recall performance at high levels of demand.

Subjects reported that they preferred to concentrate on maintaining the items in short-term memory by rehearsing the key elements which indicates they were not confident of retrieval or differentiating the previous from the current information. Many of the subjects expressed the view that they were uncertain about the quality of their responses and frequently guessed responses in the collision avoidance dialogues used in the probed recall task.

4.4 Speech recognition errors and poor performance

The poorer performance of the speech recogniser at higher demand levels for particular subjects may reflect their inability to maintain a suitably stable pattern because voice patterns can change with the increasing stress (Baber and Noyes, 1996). This possibility has been recognised in previous research, where external factors degrading speech recognition and internal responses to those factors have been recognised as significant in terms of usability criteria. Interestingly the frequency of speech errors found in this research is not too dissimilar to that reported at a recent conference which examined the current systems (Steeneken and Pijpers, 1996). Indeed, the performance of the speech recogniser with numbers used to change altitude, radio frequency, and bearing is surprising given reports of problems with digits (South, 1996). Both (Steeneken and Pijpers, 1996) and South (1996) found that voice recognition performance varied with the flight status and moderate g-levels resulted in poorer performance. Thus, adequate performance with air-borne voice recognition systems is only likely to occur in straight and level flight, this further restricts the utility of such systems in aerospace applications.

4.4 Subjective Experience of Workload and Performance

The significantly higher error rates in the high demand conditions with speech present, with or without probed recall, is suggestive of a possible cost associated with speech dialogues. Subjects' verbal reports of these conditions certainly supported this interpretation of the results. Many of the subjects had reported difficulties in maintaining an adequate model of the flight parameters in the high demand conditions, and their feeling was that the workload increased when they were required to maintain flight parameters during the probed recall version of the task. The difficulties in maintaining the flight parameters in the high demand condition were underlined by the generally poorer quality of recall in the high demand conditions and subjects more frequently mis-reported more than a single digit incorrectly in these conditions. Although performance, in the recall of flight parameters, was improved in the probed recall task, this was certainly experienced as a much greater demand on resources by the subjects. In practice, a pilot experiencing this pressure may more frequently check visual instruments in a manner that could negate one of the advantages of speech-based interaction. Constant visual checking would reduce attention in head-up modes of operation or result in the same frequency of head-down consistency checks.

Although the error rates for the probed recall tasks were small in experimental terms these could represent significant problems in real-time supervisory control tasks and the poorer quality of recall in high demand situations could provide an opportunity for latent errors to appear in long term operation of such systems. Poor recall would simply result in the type of error in decision making which has been identified as a contributory factor in aerospace accidents and in disasters with

complex supervisory control systems.

A natural counter to these suggested problems is the use of extended training and this is currently under investigation with this series of tasks. Training could be an expensive option and it may never be completely fault-tolerant in that people may revert to preferred modes in the event of an unpredictable or high demands. Another issue which remains unresolved is with regard to task design and the integration of information from diverse multi-modal systems interfaces to produce a big picture.

In conclusion, the usability of speech-based interfaces used as part of a multi-modal interface may not improve with improvements in speech recognition rates. The performance of operators using a extended multi-modal systems interface for supervisory control must be carefully designed to integrate the speech-based interfaces into the overall tasks performed and ensure that the use of speech does not increase the demands on the operator. At moderate and particularly high levels of demand there may be significant performance degradation across tasks as a consequence of selecting and designing a multi-modal interference. It is important to note that in supervisory tasks in cockpits, power plants and chemical processes the levels of demand used in this experimental series would be probably be considered moderate. Interestingly, it is these moderate levels of demand which have been recommended as the appropriate context in which to introduce speech-based systems (Cresswell-Starr, 1993).

Acknowledgements

The authors wish to acknowledge funding from the European Social Fund initiative which allowed Charles Cranmer and Carol-Ann Milton to take part in the research project as part of an advanced training technology initiative. The authors also wish to thank NCR (Scotland Ltd.) for the use of the Dialogue Designer rapid prototyping software developed under their sponsorship. The authors would also like to thank Mrs Sheila M. Whyte who provided voice segments for the simulated on-board computer.

References

Achille L.B., Schulze K.G. and Schmidt-Nielsen A. (1995) An Analysis of Communication and the Use of Military Terms in Navy Team Training, Military Psychology, 7, 2, 95-107.

Adams M.J., Tenney Y.J. and Pew R. W. (1991) Strategic Workload and the Cognitive Management of Advanced Multi-task Systems, Crew Systems Ergonomics Information Analysis Center, Wright Patterson Airforce Base, OHIO, USA.

Baber C. and Noyes J. (1996) Automatic Speech Recognition in Adverse Environments, Human Factors, 38, 1, 142-155.

Baddeley A. (1994) Human Memory : Theory and Practice, Lawrence Erlbaum Associates Ltd., Hove, United Kingdom.

Cheung B. , Money K. and Sarkar P. (1996) Loss of Aviation Situation Awareness in the Canadian Air Forces, in Situation Awareness : Limitations and Enhancement in the Aviation Environment, AGARD Symposium, AGARD-CP-575, AGARD, Neully-sure-Seine, France.

Cook M.J. and Elder L. (1996) Agent Based Multimedia Dialogues for Reduced Workload and Increased Situational Awareness in the Cockpit, in AGARD Conference Proceedings 575, Situation Awareness : Limitations and enhancement in the aviation environment, AGARD-CP-575, AGARD, Neully-sure-Seine, France.

Cook M.J., Cranmer C.J., Finan R., Sapeluk A., and Milton C.A., (1996) Latent errors in situational awareness resulting from speech-based interfaces for cockpits, paper presented to the Symposium on Situational Awareness in the Tactical Aerospace Environment at Patuxent River, Maryland, USA to be published by CSERIAC Program Office, Wright Patterson Air Force Base Ohio, USA.

Cresswell-Starr A. F. (1993) Is control by voice the right answer for the avionics environment ?, in Interactive Speech

Technology : Human Factors Issues in the application of Speech Input/Output to Computers, C. Baber and J. M. Noyes Eds., Taylor Francis, London, England.

Damper R. I. and Wood S. D. (1995) Speech versus keying in command and control applications, International Journal of Human-Computer Studies, 42, 289-305.

Damper R. I. , Tranchant M. A. and Lewis S. M. (1996) Speech versus keying in command and control: Effect of concurrent tasking, submitted to International Journal of Human-Computer Studies.

Finan R. (1994a) Dialogue Designer 1.0 : Final Report, A.T. and T. (GIS) Technical Report, A.T. and T. (GIS), Dundee.

Finan R. (1994b) Dialogue Designer 1.0 : User Manual, A.T. and T. (GIS) Technical Report, A.T. and T. (GIS), Dundee.

Finan R., Cook M.J., and Sapeluk A. (1996) Speech-based dialogue design for usability assessment, in C5 (Human-Computer Interaction) Digest No.: 96/126.

Gerlach M. and Onken R. (1993) A Dialogue manager as an interface between aircraft pilots and a pilot assistant system, in Human Computer Interaction, Proceedings of Fifth International Conference on HCI (HCI International '93), Salvendy G. and Smith M.J. Eds., 98-103, Volume 2.

Onken R. (1995) Human-centred cockpit design through the knowledge-based cockpit assistant system CASSY, paper presented at AGARD Symposium on Situational Awareness, DRG Panel 8, Technical Proceedings AC/243 (Panel 8), TP/7, AGARD/NATO Publication.

Raymond R., and Isaac A.R. (1995) Simultaneous error - during altitude deviations in Aviation Psychology: Training and Selection, Johnston N., Fuller R., and McDonald N. Eds., Avebury Aviation, Ashgate Publishing Company, Aldershot, Hants., England.

Rudnicky A.I. and Hauptmann A.G. (1992) Multimodal Interaction In Speech Systems, in Multimedia Interface Design. Edited by M.M. Blattner and R.B. Dannenberg, ACM Frontier Series, Addison-Wesley, Wokingham.

Steeneken H.J.M. and Pijpers E.W. (1996) Development and Performance of Cockpit Control Systems Operated by Voice, paper presented at AGARD AMP Special Meeting on Audio Effectiveness in Aviation to be published as conference proceedings by AGARD, Neuilly-sur Seine, France.

Steeneken H.J.M. and Gagnon L. (1996) Coordinated Speech Technology Research of Nine NATO Countries in Research Study Group RSG 1.0,paper presented at AGARD AMP Special Meeting on Audio Effectiveness in Aviation to be published as conference proceedings by AGARD, Neuilly-sur Seine, France.

South A.J. (1996) Voice Recognition in Adverse Cockpit Environments, paper presented at AGARD AMP Special Meeting on Audio Effectiveness in Aviation to be published as conference proceedings by AGARD, Neuilly-sur Seine, France.

Stokes A. , Wickens C. and Kite K. (1990) Display Technology : Human Factors Concepts, Society of Automotive Engineers, Warrendale, PA, USA.

Turner J. (1995) Personal communication with EFA Chief Test Pilot at AGARD Symposium on Situation Awareness : Limitations and enhancement in the aviation environment, Brussels.

Turner J. (1996) Personal communication with EFA Chief Test Pilot at AGARD Symposium on Selection and Training Advances in Aviation, Prague.

Wickens C. D. (1984) Processing resources in attention, in Varieties of Attention (pp. 63-101), Parasuraman R. and Davies R. Eds. , Academic Press, London.

Wickens C. D. (1989) Attention and skilled performance, in Human Skills, 2nd Edition, Holding D. Ed. , John Wiley and Sons, Chichester, England.

Wickens C. D. (1992) Engineering Psychology and Human Performance 2nd Editon, Harper and Collins Publishers Inc. , New York, USA.

Wickens C. D. and Flach J. (1988) Human Information Processing. in Human Factors in Aviation, New York, Academic Press, London.

Wickens C. D. , Sandry D. L. and Vidiluch M. (1983) Comaptability and resource competition between modailities of input, central processing and output, Human Factors, 25, 2, 227-248.

Wickens C. D. , Vidiluch M. and Sandry-Garza D. L. (1984) Principles of S-C. R compatability with spatial and verbal tasks : The role of display-control location and voice-interactive display-control interfacing, Human Factors, 26, 5, 533-543.

Woods D.D., Johannesen L.J., Cook R.I. and Sarter N.B. (1995), Behind Human Error : Cognitive Systems, Computers, and Hindsight. Crew Systems Ergonomics Information Analysis Center, Wright Patterson Airforce Base, Ohio, USA.

Zue V.W., Glass J.R., Goddeau D., Goodine D., Hirschmann L., Leung H.C., Phillips M.S., Polifoni J., and Seneff S. (1991) Spoken Language Systems for Human/Machine Interfaces, 926-955, in Procs. Intelligent Text and Image Handling (RIAO '91), Elsevier Press, Amsterdam.

Graph 1 : Mean percentage misses and standard errors plotted for low, medium and high demand tasks with and without speech.

Figure 1 : Low demand squares task.

Figure 2 : Moderate demand squares task.

Figure 3 : High demand squares task.

Figure 4: Engine Status Dialogue Structure.

Figure 5 : Altitude Dialogue Structure.

Figure 6 : Collision Avoidance Dialogue Structure.

Figure 7 : Initial Flight Parameter Recall Performance.

Figure 8 : Final Flight Parameter Recall Performance.

Figure 9 : Additional Flight Parameter Recall Performance.

Figure 10 : Initial Flight Parameter Recall With and Without Probed Recall.

Figure 11 : Final Flight Parameter Recall With and Without Probed Recall.

Figure 12 : Number of Incorrect Responses for First Probe Dialogue Set.

Figure 13 : Number of Incorrect Responses for Second Probe Dialogue Set.

Level of Demand (Workload)	**Without Speech Tasks** (Control Condition)	**With Speech Tasks** (Experimental Condition)
Low	Arrow (Orientation) Task Square(Colour) Task	Arrow (Orientation) Task Square(Colour) Task Speech Dialogues
Medium	Arrow (Orientation) Task Square(Easy Discrimation)	Arrow (Orientation) Task Square(Easy Discrimination) Speech Dialogues
High	Arrow (Orientation) Task Square(Hard Discrimination)	Arrow (Orientation) Task Square(Hard Discrimination) Speech Dialogues

Table 1 : Six conditions examined in the experimental series.

Level of Demand (Workload)	**Without Speech Tasks** (Control Condition)	**With Speech Tasks** (Experimental Condition)
Easy Square Task **Low**	Arrow 3.1 ± 3.4 Square 2.3 ± 3.4	Arrow 11.2 ± 10.8 Square 4.5 ± 5.0
Moderate Square Task **Medium**	Arrow 4.2 ± 6.8 Square 8.1 ± 5.1	Arrow 4.9 ± 5.9 Square 13.8 ± 10.8
Hard Square Task **High**	Arrow 11.7 ± 12.7 Square 35.4 ± 13.7	Arrow 5.6 ± 6.8 Square 58.5 ± 18.0

Table 2: Mean percentage of misses and standard deviations for the arrows and squares tasks at three levels of workload.

Level of Demand (Workload)	**Without Speech Tasks** (Control Condition)	**With Speech Tasks** (Experimental Condition)
Easy Square Task **Low**	Arrow 0.1 ± 0.3 Square 0.2 ± 0.4	Arrow 0.6 ± 1.6 Square 0.6 ± 1.1
Moderate Square Task **Medium**	Arrow 0.2 ± 0.4 Square 0.6 ± 1.0	Arrow 0.3 ± 0.7 Square 0.2 ± 0.6
Hard Square Task **High**	Arrow 0.6 ± 0.7 Square 5.0 ± 6.3	Arrow 0.5 ± 0.7 Square 3.3 ± 2.6

Table 3 : Mean percentage of false alarms for arrows and squares tasks for 3 levels of workload.

Level of Demand (Workload)	**Without Speech Tasks** (Control Condition)	**With Speech Tasks** (Experimental Condition)
Easy Square Task **Low**	Arrow 3.1 (4) Square 2.3 (4.5)	Arrow 11.2 (3.5) Square 4.5 (3.5)
Moderate Square Task **Medium**	Arrow 4.2 (4) Square 8.1 (3.5)	Arrow 4.9 (4) Square 13.8 (3)
Hard Square Task **High**	Arrow 11.7 (3) Square 35.4 (2)	Arrow 5.6 (1) Square 58.5 (1)

Table 4 : Comparison of median scores for subjective performance assessment (in brackets)[1] and mean percentage of misses on each task at three levels of demand.

[1] Performance was assessed using a 5-point scale, where 1 = very poor and 5 = very good.

Workload	Speech Recognised	Speech Unrecognised	Speech Misrecognised	Hesitation[2] Error
Low	75.5	14.5	8.0	2.0
Medium	78.5	8.5	6.5	6.5
High	66.0	26.0	5.5	2.5

Table 5 : Mean speech recognition rates for low, medium and high demand conditions.

	Condition		
Dialogue Type	Low	Medium	High
Altitude	4	4	4
Bearing	3	3	4
Frequency	2.5	3	3
Engine Status	3	4	3
Undercarriage Status	4	4	4
Combined Dialogue Altititude and Bearing	2	2	2
Combined Dialogue Undercarriage and Bearing	2	3	2

Table 6 : Median of subjective assessment of performance for different dialogues at low, medium, and high demand.

[2] Hesitation occurred when the subject failed to provide a response within a given period of time.

Recall Error	Low Demand	Medium Demand	High Demand
No answer given	39.3	31.6	29
One digit incorrect	46.4	21.1	17
More than one digit incorrect.	14.3	47.3	54
Total (with no answer / more than one digit incorrect)	53.6	78.9	83.0

Table 7 : Type of flight parameter recall error for low, medium, and high demand conditions.

Workload and Task Combinations				
	No Probed Recall Task		**Probed Recall Task**	
	High Demand Squares Task + Arrow Task	High Demand Squares Task + Arrow Task + Speech Tasks	High Demand Squares Task + Arrow Task	High Demand Squares Task + Arrow Task + Speech Tasks
Arrow Task Misses	11.7 ± 12.7	5.6 ± 6.8	5.2 ± 4.1	5.5 ± 4.8
Squares Task Misses	35.4 ± 13.7	58.5 ± 18.0	28.7 ± 10.9	60.4 ± 17.2

Table 8 : Mean percentage misses and standard deviations for arrows and squares tasks for high workload condition with and without probed recall.

Workload and Task Combinations				
	No Probed Recall Task		Probed Recall Task	
	High Demand Squares Task + Arrow Task	High Demand Squares Task + Arrow Task + Speech Tasks	High Demand Squares Task + Arrow Task	High Demand Squares Task + Arrow Task + Speech Tasks
Arrow Task False Alarms	0.6 ± 0.7	0.5 ± 0.7	0.0 ± 0.0	0.4 ± 0.5
Squares Task False Alarms	5.0 ± 6.3	3.3 ± 2.6	4.0 ± 4.1	3.6 ± 4.1

Table 9 : Mean percentage of false alarms and standard deviations for arrow and squares tasks under high workload with and without probed recall.

Workload and Task Combinations				
	No Probed Recall Task		Probed Recall Task	
	High Demand Squares Task + Arrow Task	High Demand Squares Task + Arrow Task + Speech Tasks	High Demand Squares Task + Arrow Task	High Demand Squares Task + Arrow Task + Speech Tasks
Arrow Task False Alarms	11.7 **(3)**	5.6 **(1)**	5.2 **(2)**	5.5 **(3)**
Squares Task False Alarms	35.4 **(2)**	58.5 **(1)**	28.7 **(2)**	60.4 **(1)**

Table 10 : Comparison of median scores for subjective performance assesment [3](shown in bold lettering) and mean percentage error scores for visual tasks in high demand condition with and without probed recall.

[3] Performance was assessed using a 5-point scale, where 1 = very poor and 5 = very good.

Workload	Speech Recognised	Speech Unrecognised	Speech Misrecognised	Hesitation Error
High Demand and No Probed Recall	66.0	26.0	5.5	2.5
High Demand and Probed Recall	70.6	20.0	5.0	4.4

Table 11 : Mean speech recognition rates for high demand conditions with and without probed recall.

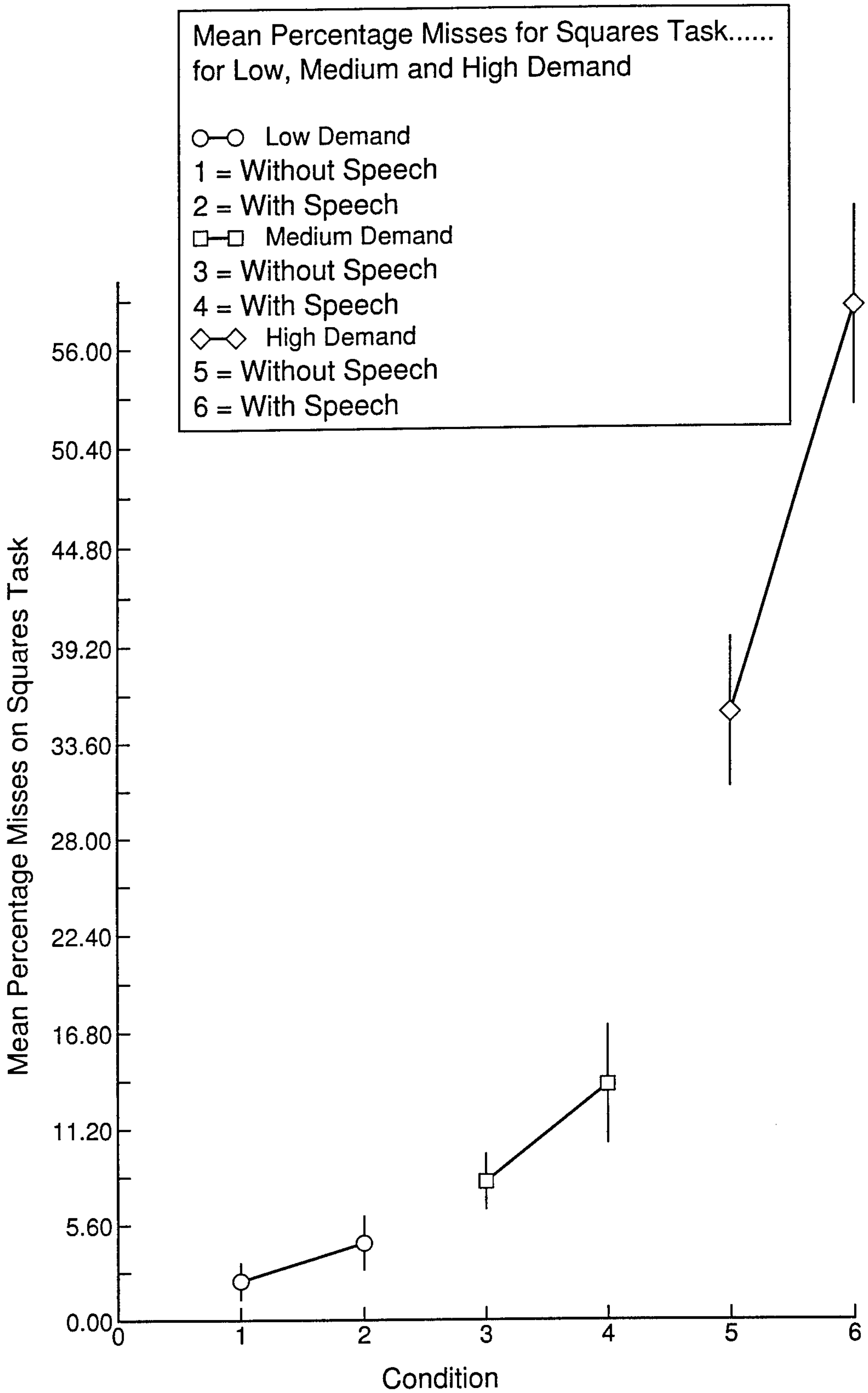
	Condition	
Question Type ↓	**High Demand Task**	**High Demand Task and Probed Recall**
Altitude	4	3.5
Bearing	4	4
Frequency	3	3.5
Engine Status	3	3
Undercariage Status	4	N/A
Combined Dialogue Altititude and Bearing	2	N/A
Combined Dialogue U/carriage and Bearing	2	N/A

Table 12 : Median Scores [4] for dialogue difficulty for high demand with and without probed recall.

[4] Performance was assessed using a 5-point scale, where 1 = very difficult, and 5 = very easy. N/A indicates that this question was not part of the questionnaire in this condition.

	Final Parameters		
Condition	Any engines shutdown ? (% correct answers)	How many engines shutdown ? (% correct answers)	Which engine number shutdown ? (% correct answers)
High Demand	100	100	100
High Demand and probed recall task	100	100	100

Table 13 : Additional flight parameter recall - mean percentage correct.



Mean % Misses on Squares Task for 3 Demand Levels

Graph 1 : Mean percentage misses and standard errors plotted for low, medium and high demand tasks with and without speech.

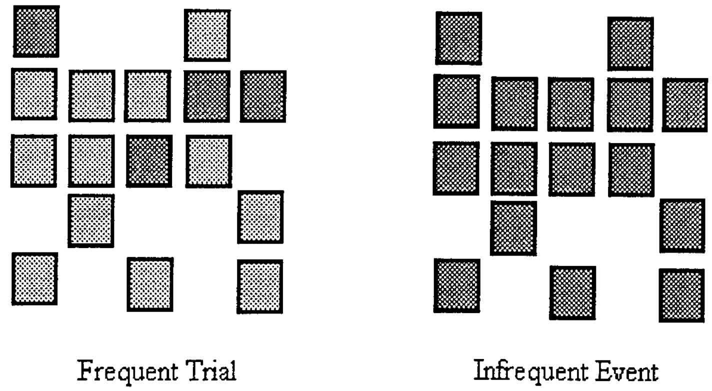


Figure 1 : Low demand squares task.

Low Demand Squares Task : Pattern repeats on every trial and on the significant event trial all the squares turn red (indicated by darker squares).

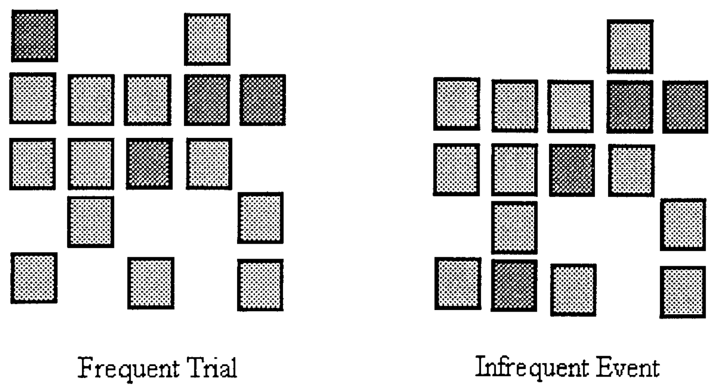


Figure 2 : Moderate demand squares task.

Moderate Demand Squares Task : Pattern repeats on every trial and on the significant events a single red square moves.

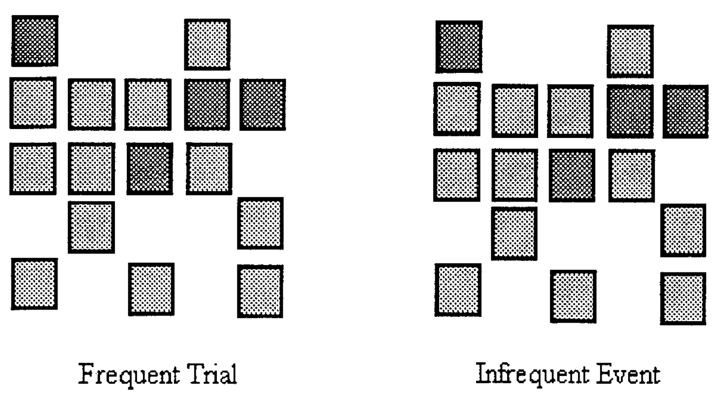


Figure 3 : High demand squares task.

High Demand Squares Task : Pattern changes on every trial except in the significant event when the pattern repeats.

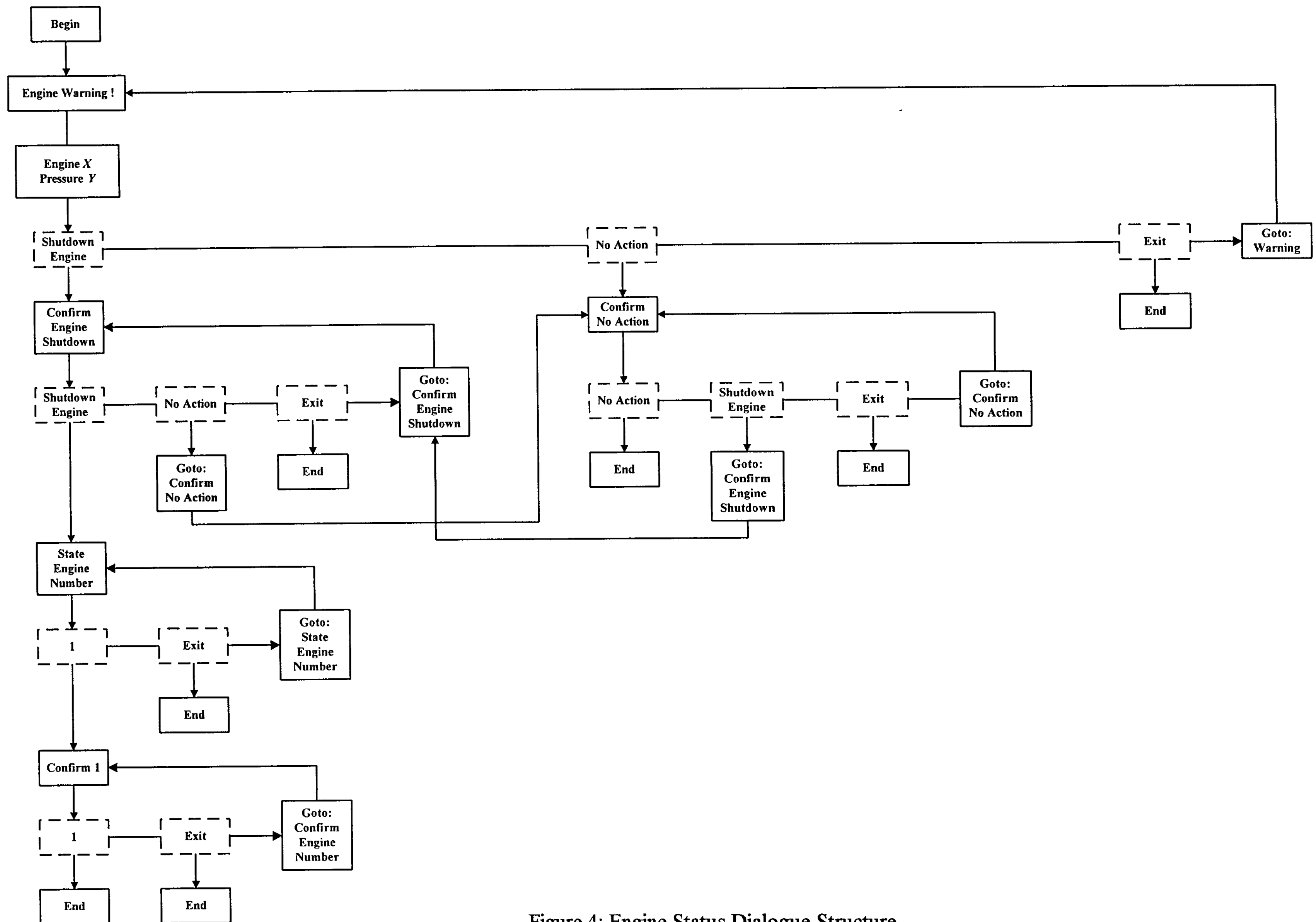


Figure 4: Engine Status Dialogue Structure.

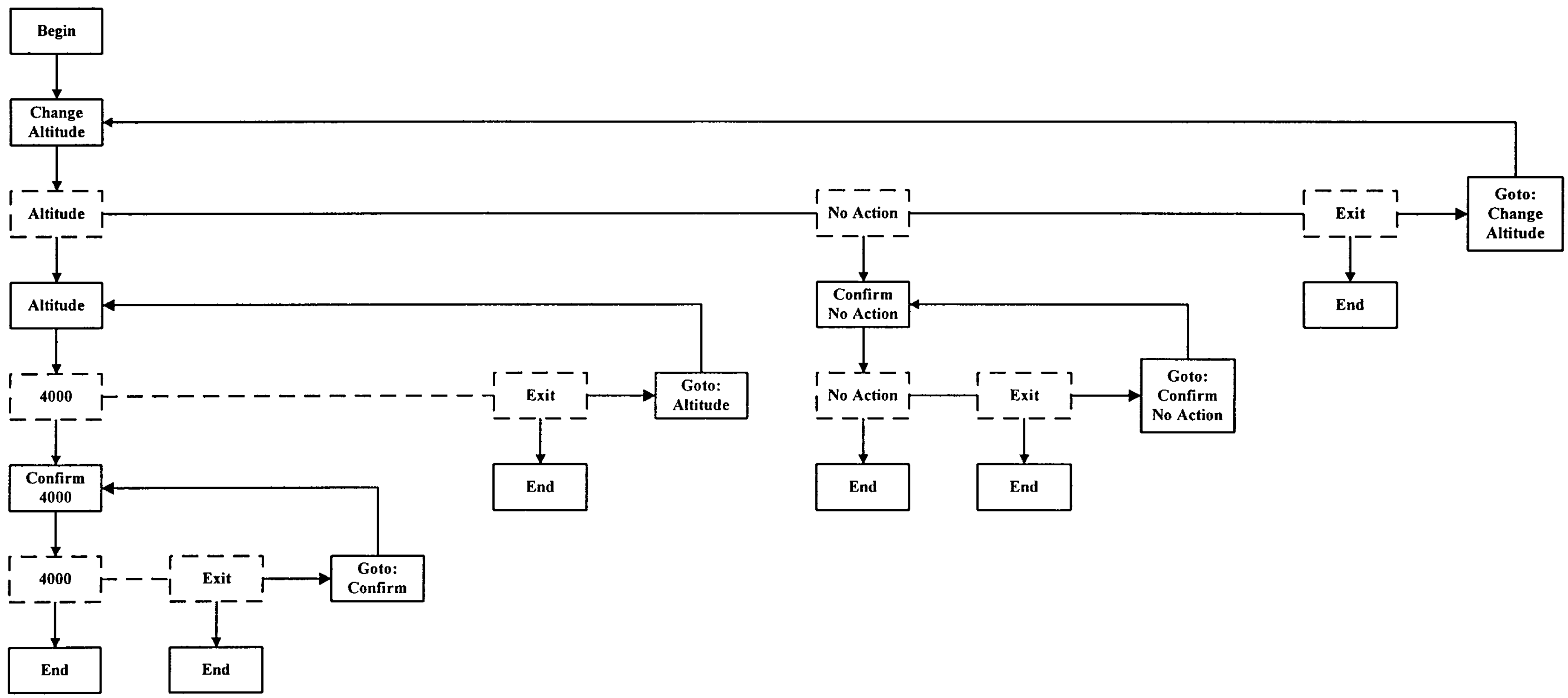


Figure 5 : Altitude Dialogue Structure.

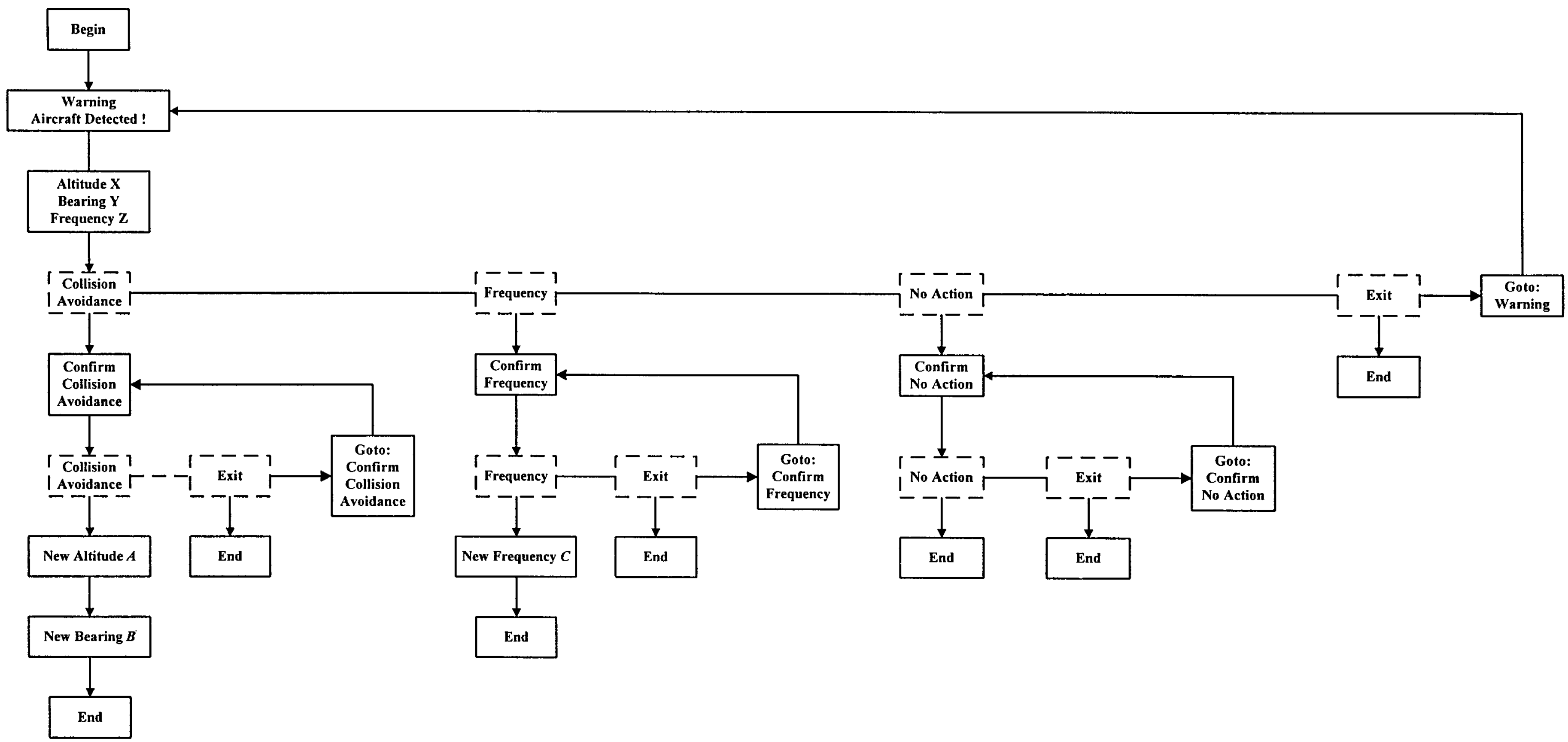


Figure 6 : Collision Avoidance Dialogue Structure.

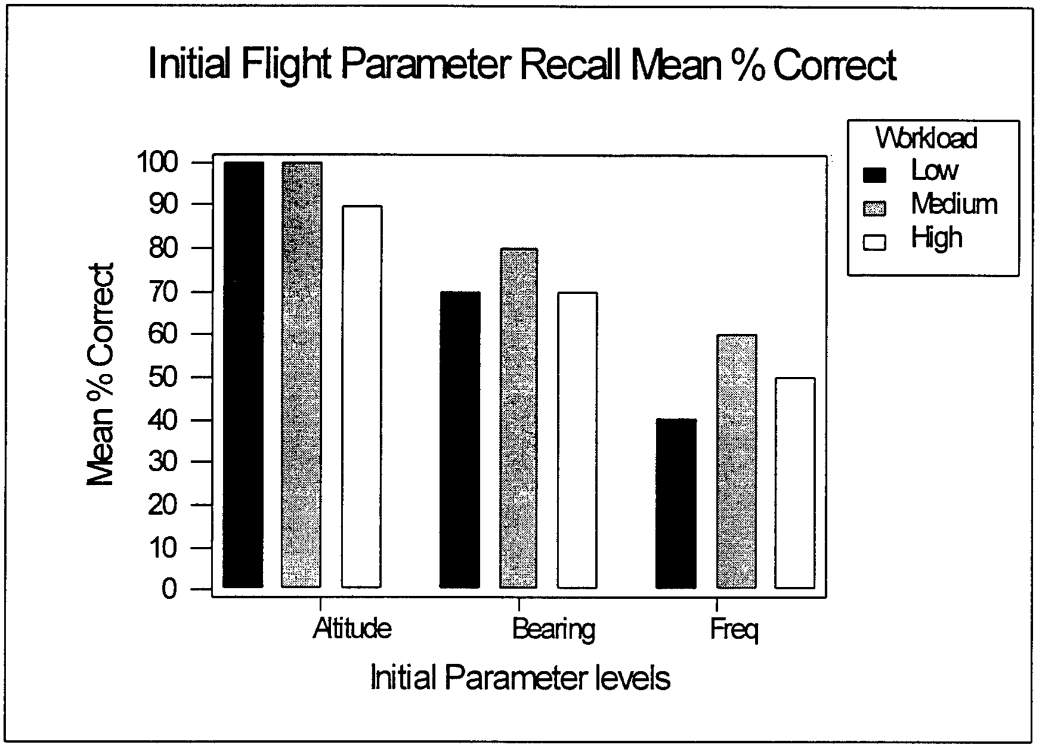


Figure 7 : Initial Flight Parameter Recall Performance.

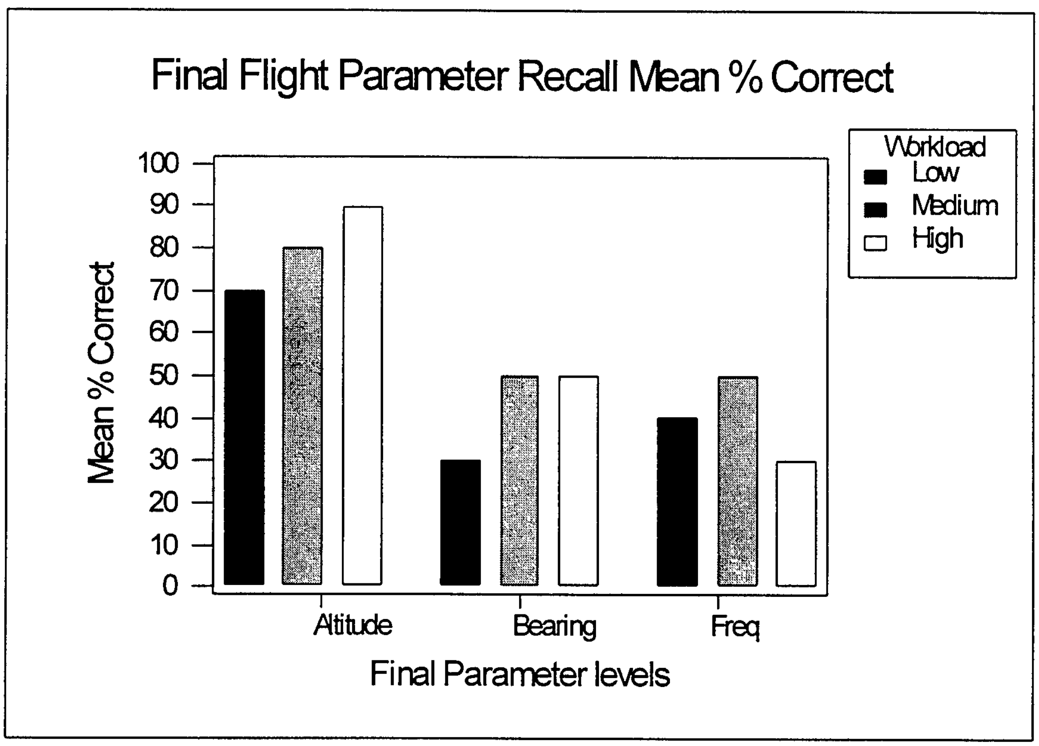


Figure 8 : Final Flight Parameter Recall Performance.

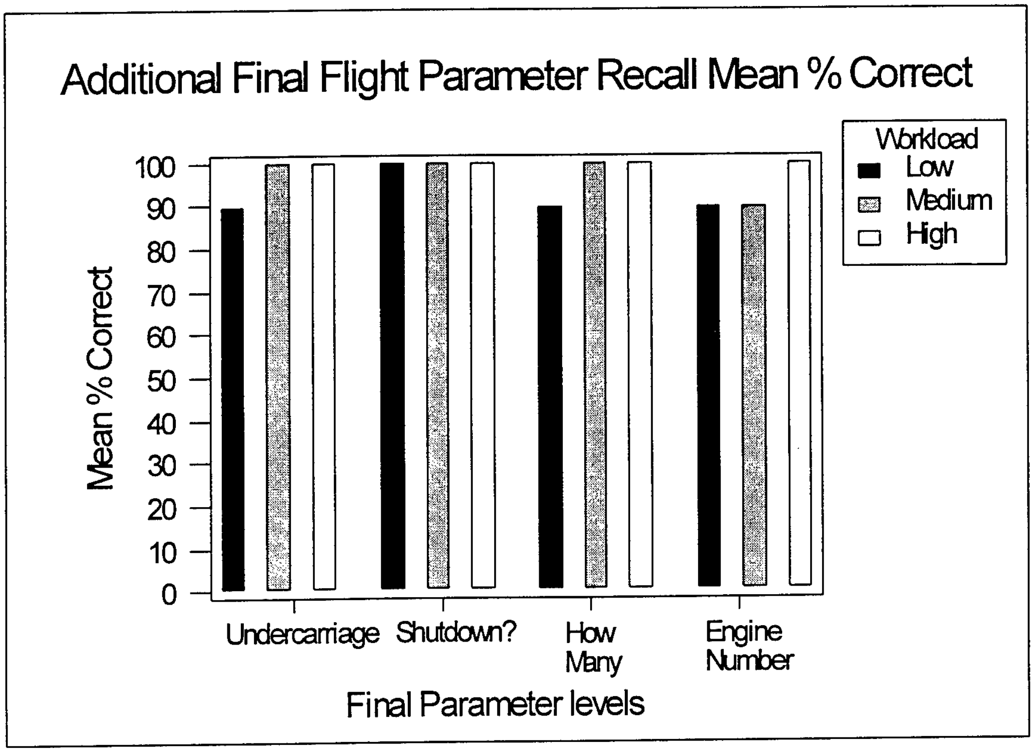


Figure 9 : Additional Flight Parameter Recall Performance.

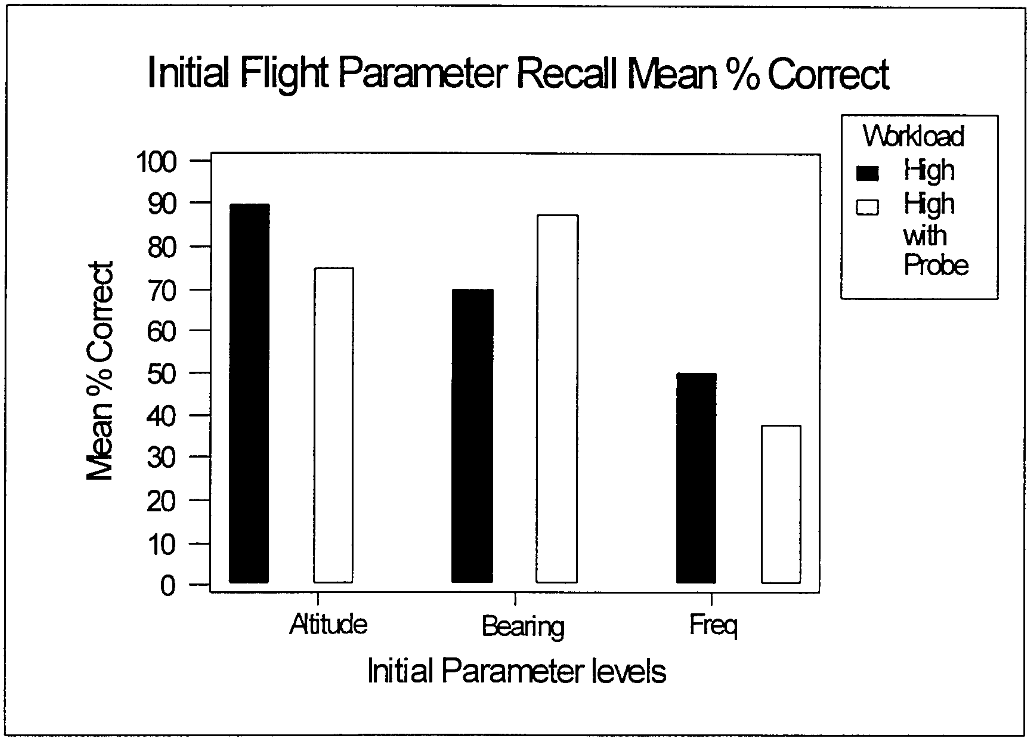


Figure 10 : Initial Flight Parameter Recall With and Without Probed Recall.

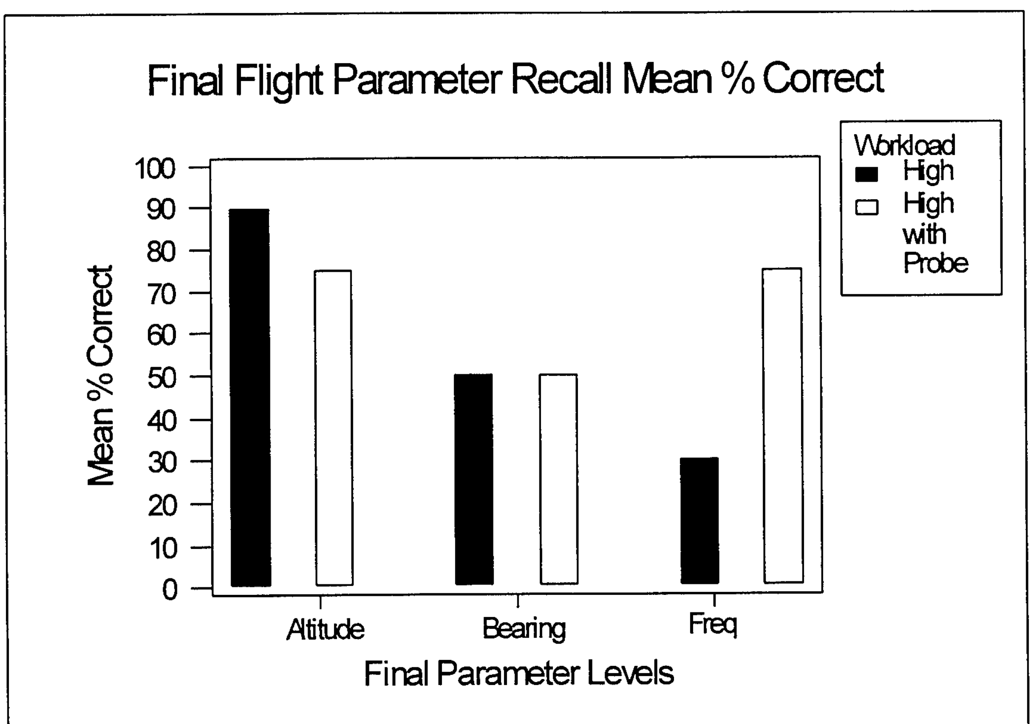


Figure 11 : Final Flight Parameter Recall With and Without Probed Recall.

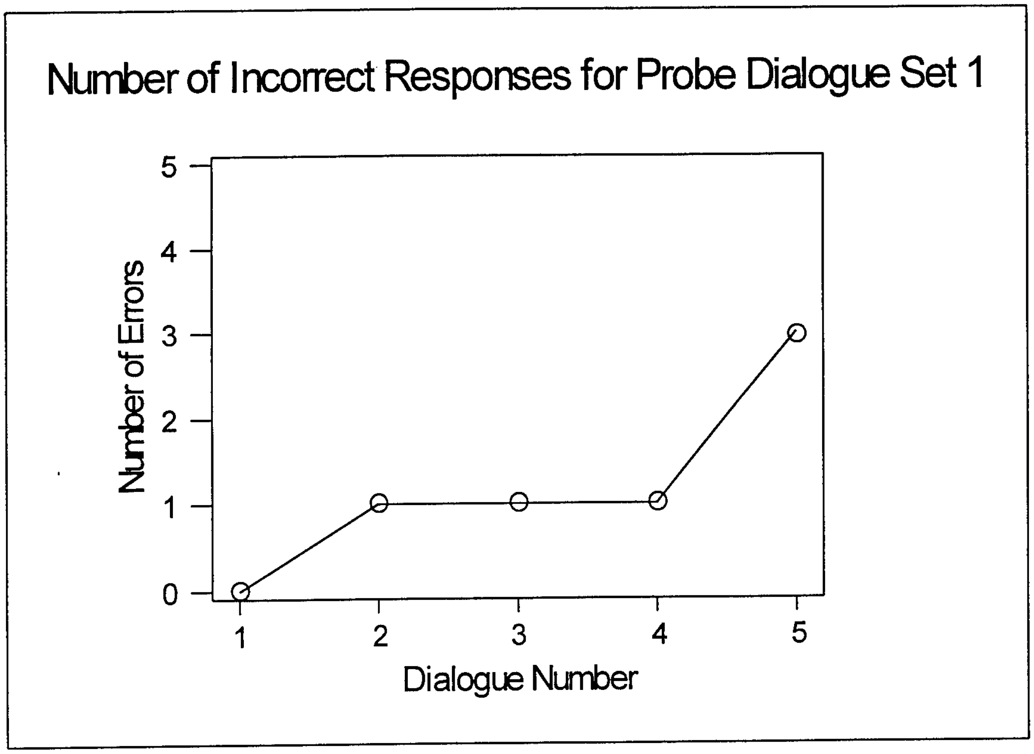


Figure 12 : Number of Incorrect Responses for First Probe Dialogue Set.

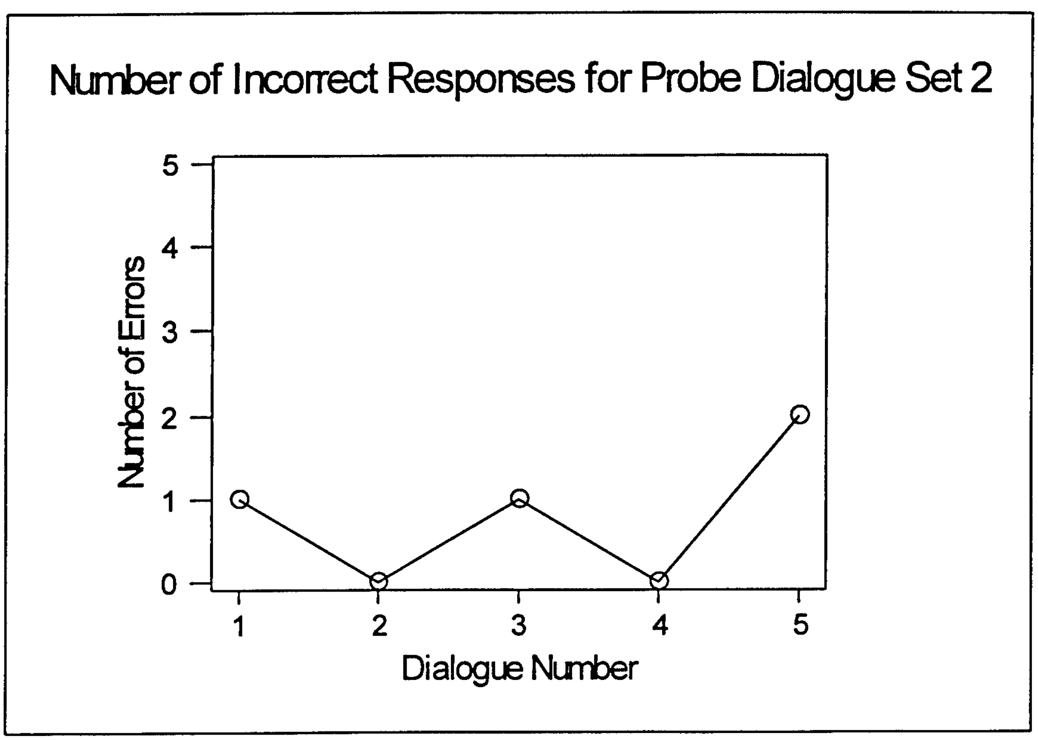


Figure 13 : Number of Incorrect Responses for Second Probe Dialogue Set.

SUMMARY OF THE MEETING ON AUDIO EFFECTIVENESS IN AVIATION

Thomas J. Moore

Crew Systems Directorate
Armstrong Laboratory, AL/CFBA
Wright-Patterson AFB OH 45433-7901, USA

Mister Chairman, ladies, and gentlemen

I am very pleased to have the opportunity to make some summary comments on the excellent meeting that we are now concluding. The topic of Audio Effectiveness in military and civil aviation is an important one and one that too seldom is given the attention it deserves. The last time this organization specifically addressed this topic was in a conference on "Aural Communications in Aviation" held in Soesterberg, Netherlands, 15 and 1/2 years ago.

At that conference there was a total of 25 papers presented, the overwhelming majority of which dealt with the effects of the aircraft noise environment. The effects discussed were primarily Noise Induced Hearing Loss in Aviators and the intelligibility of voice communications in noisy environments.

Over 15 years later, the aircraft noise environment is still one of our major concerns, as evidenced by the keynote address by Dr. Rood. This concern is not only still with us, but in many cases the noise environment in which aviators operate is becoming even more severe. This is due in large part to the desire to minimize aircraft weight wherever possible, leading to less sound attenuation treatment and the increased use of composite materials. The papers presented in the Noise Control session of this conference on Tuesday focused on the use of new materials and technologies to reduce the level of noise at the operators' ears and thereby mitigate the chances of causing physiological damage to the auditory system, as well as effectively increasing the signal-to-noise ratio of the audio signal.

Two papers reported on the use of earplugs, equipped with audio transducers, to replace the earcups found in standard helmets. While some potential problems were identified and solutions proposed for the use of this technology in high performance aircraft, its use was deemed acceptable in rotary-winged aircraft, leading to extended discussions that will continue long after this meeting concludes.

Particularly striking is the growth in interest in the use and application of Active Noise Reduction (ANR) Technology. At the Soesterberg conference, there was one paper on this technology. At the present conference, there are 12 papers dealing with the evaluation and application of ANR systems.

A number of these papers provided comparative evaluations of the performance of commercially available systems. These evaluations, while providing a reasonably consistent picture of the advantages and shortcomings of different systems. Also illustrated, as Dr. Steenekin noted, the need

Paper presented at the AMP Symposium on "Audio Effectiveness in Aviation", held in Copenhagen, Denmark, 7-10 October 1996, and published in CP-596.

for the development of a standardized methodology for the evaluation of ANR systems.

Other papers dealt with the application of ANR technology in helicopter, fighter jet and armored vehicle environments. We also had some discussion of future development of this technology. Among these were the development of ANR integrated into earplugs, allowing use with chemical defense ensembles, the development of hybrid analog/digital systems that will allow greater attenuation, faster response times and adaptive control, and the use of the technology to enhance the performance of hearing impaired individuals in noise environments.

Overall, the presentations at this conference demonstrate that ANR is a technology that has matured sufficiently that its application in operational environments is underway, even while there are problems still to be addressed and further development to be accomplished.

A second major portion of this conference was devoted on Monday to the topic of Audio Displays. These papers were all concerned with the ability to provide spatial auditory information to the operator and thereby enhance performance in the cockpit. The work presented ranged from basic to applied. We heard papers dealing with research on the ability to localize auditory signals in noise, the optimization of audio signals to enhance distinctiveness and localization, the interaction of audio and visual signals, and the utility of "3-D" audio cues in simulator studies and flight demonstrations.

The consensus emerging from these presentations is that the use of spatial audio cues provides a clear synergistic effect with the use of two-dimensional visual cues for target detection, significantly enhances the intelligibility of communications, and promises increased situation awareness for the operator. It is clear that the presentation of spatial audio information over earphones, which was not thought feasible 15 years ago, holds great promise for future cockpit applications.

The final area with which this conference dealt was that of Speech Technology. Here again we see that the question of the aircraft noise environment and its effects on voice communications, is still a problem with which our community is concerned. One important question that was addressed in this session was how to measure the effectiveness of voice communications. In other words, how do we measure the amount of information that is being communicated within the constraints of a specific operational scenario, rather than what percentage of a list of words is correctly identified. This is an important question and deserves further research.

In the speech technology session there were also a number of papers that addressed the application of automatic speech recognition (ASR) technology in aviation and the difficulties experienced with this technology in the flight environment. These difficulties are attributable to the environmental stressors experienced in flight (noise, vibration, acceleration) which not only exert varying influences on the speech produced by the operator (e.g., increased vocal effort, voice tremor), but also directly affect the performance of the recognition system (e.g., noise obscuring word boundaries which would affect recognizers based on template matching techniques). These environmental factors and others, such as

emotional stress, all cause acoustic-phonetic changes in the speech signal that can and do influence the performance of the recognition system. Finally, the equipment that serves as the means through which the speech is transformed into an electrical signal that is the input to the ASR device is an important element in determining the success or failure of the device (e.g., the microphone, oxygen mask and audio distribution system in the case of aircraft).

For the near future, limited vocabulary, speaker dependent ASR systems are the only ones that appear viable for cockpit applications. Even given these constraints, there are sufficient data entry and information retrieval applications that automatic speech recognition technology promises to be a valuable tool for reducing workload and enhancing efficiency of the pilot/operator.

I would like now to briefly mention some topics that were not fully considered at this conference.

The first topic is the issue of intelligibility of speech in noise for non-native speakers of the language. That is, how are communications affected if the communications are in a language which is not the native language of one of the communicators. At this meeting, this topic was mentioned by Dr. Steenekin in his overview of the activities of RSG 10 and Dr. Hanschke in his paper on audiometric standards for aviators. This is a question that has important implications, particularly in the case of a multi-national organization such as NATO. At the Soesterberg meeting, data were presented where simple Norwegian and English sentences read by a bilingual speaker were recorded and presented embedded in noise to bilingual listeners who had either Norwegian or English as their first language and had a good command of the second language. The results were that for both language groups the native language sentences were correctly perceived at a lower signal-to-noise ratio than were the non-native language sentences. Other studies have shown that non-native speakers perform poorer than native speakers when listening to synthetic or digitally encoded speech and showed a greater degradation of intelligibility when these signals were presented in the presence of noise.

This is an area of research that needs more study and should be of interest to a great number of people at this meeting.

Another matter that should concern many of us is the question of acceptance by the operators of the new technologies we develop. Often with some justification the operator is reluctant to embrace the latest product from the laboratories either because there is a perception that the function or operation that we are attempting to improve works just fine as it stands, or that our attempts to solve various problems will only create new ones. If, however, attention is paid to a number of elements during the course of the development of a technology, the probability of a successful transition from the laboratory to the field is greatly enhanced. Among these elements are: (1) early identification of the customer--the sooner the customer/user can identified the more likely it is that when the technology is ready there will be a smooth transition into the next stage of development or incorporation into existing or planned programs; (2) involvement of operators during the development--once the customer has been identified, representatives of the types of operators who will be using the systems, e.g. pilots, communicators, intelligence analysts, etc., should be consulted

during the development process. Often insights provided by these experienced operators will influence the design of the technology. In many cases, inputs from the operators at this stage will greatly increase the acceptance of the technology by the operational community when it is finally ready to be fielded; (3) be prepared to market your product--as a laboratory scientist/engineer concerned with the eventual successful application of technology that you helped develop, you have to be willing to help "market" the product and not just "ship it out the door." One of the best ways to do this is to develop demonstration models of your technology that show its' capabilities and potential advantages to the operational community. At a minimum this should be a laboratory based demonstration so that potential "customers" visiting your laboratory can be exposed to the technology under development. Even better is if the demonstration can be packaged so that it is capable of being taken on the road to professional meetings and operational sites where a cross section of potential users can be exposed to it. A good demonstration is often the best way to generate a user requirement for your technology, since in many cases the user may be unaware that there is a better way to accomplish the task at hand.

In our laboratory an essential contributor to the successful acceptance by the operator of the ANR technology has been an audio demonstration booth we fabricated in-house. During the development of the ANR system we often encountered aviators who expressed reservations about using ANR because they feared that it would reduce the level of not only unwanted noise, but also the level of communications and audio cues that they felt allowed them to maintain awareness of the status of their aircraft. Basically we took a single person audiometric test booth and modified it by installing a grid of 4 1/2-inch speakers in the ceiling. With this arrangement we were able to generate up to 130 dB Sound Pressure Level (SPL) over a frequency range from 100 Hz to 10 kHz within the booth. Using this capability we developed a demonstration where an individual seated in the booth can select the recording of the noise at the pilot's position of a helicopter (UH-IN), a high performance fighter (F-15A) , or a turbo-prop transport (C-130), with and without communications present. In these environments, the listener can, by throwing a toggle switch either activate or bypass the ANR circuitry. With this capability we were able to demonstrate the benefits of ANR in a cockpit acoustic environment to operators both in our lab and in the field.

Also, as technology is developed in the laboratory it must be remembered it will often function as one component in a system. How it will interface with other system components and how they may affect the performance of the technology you developed must be considered. A good example is the development of ASR technology for application in a cockpit. For this application, consideration has to be given to the microphones that will be used in the field, whether or not an oxygen mask will be worn, and what are the characteristics of the audio distribution system that will be aboard the aircraft. As was found in the AFTI/F-16 demonstration effort (1987), the current intercom system is not designed to meet ASR requirements (e.g., it has a band width of 3.4 kHz). In order to demonstrate the utility of ASR technology in the cockpit, it has often been necessary to provide a separate amplifier in parallel with the existing intercom system. Until a high fidelity, wide band width audio distribution system is available in the cockpit, the likelihood that ASR will become

operational in fighter aircraft is highly problematical.

Similar challenges face the field application of 3-D audio technology, about which we have heard so much at this meeting. This technology requires a high quality audio distribution system with binaural output, as well as a head-tracker to determine the operator's head position. Until these capabilities are available we will not be able to fully utilize this technology in the cockpit.

Finally, I would like to close by expressing gratitude to all the presenters and participants in this meeting. It has been most informative and enjoyable, with many lively discussions. I wish you all a safe journey home and look forward to meeting again to discuss audio technology in aviation somewhat sooner that 15 1/2 years from now.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-596	ISBN 92-836-0043-6	UNCLASSIFIED/ UNLIMITED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly-sur-Seine, France

6. Title
Audio Effectiveness in Aviation

7. Presented at/sponsored by
The Aerospace Medical Panel Symposium held in Copenhagen, Denmark, 7-10 October 1996

8. Author(s)/Editor(s)	9. Date
Multiple	June 1997

10. Author's/Editor's Address	11. Pages
Multiple	370

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other AGARD unclassified publications is given on the back cover.

13. Keywords/Descriptors

Voice communication
Aerospace medicine
Human factors engineering
Auditory perception
Active Noise Reduction (ANR)
Ear protectors
Display devices
Flight tests
Speech recognition
Man machine systems
Aircraft noise
Noise reduction
Warning systems
Military aircraft
Civil aviation
Cockpits
Stress (psychology)

14. Abstract

These proceedings include the Technical Evaluation Report, a Keynote Address, three overview addresses of key technical areas, 34 solicited papers, and a Summary paper of the Symposium sponsored by the AGARD Aerospace Medical Panel held in Copenhagen, DE, from 7-11 October 1996.

Topics addressed during this Symposium were:

- Audio Displays
- Noise Control - Passive Technique
- Noise Control - Active Technique
- Noise Control - Applications
- Communication in Stressful Environment
- Voice Control

AGARD

NATO - OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Télécopie 0(1)55.61.22.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS

AGARD NON CLASSIFIEES

Aucun stock de publications n'a existé à AGARD. A partir de 1993, AGARD détiendra un stock limité des publications associées aux cycles de conférences et cours spéciaux ainsi que les AGARDographies et les rapports des groupes de travail, organisés et publiés à partir de 1993 inclus. Les demandes de renseignements doivent être adressées à AGARD par lettre ou par fax à l'adresse indiquée ci-dessus. *Veuillez ne pas téléphoner*. La diffusion initiale de toutes les publications de l'AGARD est effectuée auprès des pays membres de l'OTAN par l'intermédiaire des centres de distribution nationaux indiqués ci-dessous. Des exemplaires supplémentaires peuvent parfois être obtenus auprès de ces centres (à l'exception des Etats-Unis). Si vous souhaitez recevoir toutes les publications de l'AGARD, ou simplement celles qui concernent certains Panels, vous pouvez demander à être inclu sur la liste d'envoi de l'un de ces centres. Les publications de l'AGARD sont en vente auprès des agences indiquées ci-dessous, sous forme de photocopie ou de microfiche.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordonnateur AGARD-VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Directeur - Gestion de l'information
(Recherche et développement) - DRDGI 3
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
Netherlands Delegation to AGARD
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
2700 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

Le centre de distribution national des Etats-Unis ne détient PAS de stocks des publications de l'AGARD.

D'éventuelles demandes de photocopies doivent être formulées directement auprès du NASA Center for AeroSpace Information (CASI) à l'adresse ci-dessous. Toute notification de changement d'adresse doit être fait également auprès de CASI.

AGENCES DE VENTE

NASA Center for AeroSpace Information
(CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
Etats-Unis

The British Library
Document Supply Division
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Les demandes de microfiches ou de photocopies de documents AGARD (y compris les demandes faites auprès du CASI) doivent comporter la dénomination AGARD, ainsi que le numéro de série d'AGARD (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Veuiller noter qu'il y a lieu de spécifier AGARD-R-nnn et AGARD-AR-nnn lors de la commande des rapports AGARD et des rapports consultatifs AGARD respectivement. Des références bibliographiques complètes ainsi que des résumés des publications AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
publié par la NASA Scientific and Technical
Information Division
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements and Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 22161
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada Inc.
(membre de la Corporation St-Joseph)
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

AGARD

NATO - OTAN

7 RUE ANCELLE • 92200 NEUILLY-SUR-SEINE

FRANCE

Telefax 0(1)55.61.22.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED

AGARD PUBLICATIONS

AGARD holds limited quantities of the publications that accompanied Lecture Series and Special Courses held in 1993 or later, and of AGARDographs and Working Group reports published from 1993 onward. For details, write or send a telefax to the address given above. *Please do not telephone.*

AGARD does not hold stocks of publications that accompanied earlier Lecture Series or Courses or of any other publications. Initial distribution of all AGARD publications is made to NATO nations through the National Distribution Centres listed below. Further copies are sometimes available from these centres (except in the United States). If you have a need to receive all AGARD publications, or just those relating to one or more specific AGARD Panels, they may be willing to include you (or your organisation) on their distribution list. AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-major de la Force aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Research & Development
Information Management - DRDIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
2700 Amadora

SPAIN
INTA (AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar-Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Center for AeroSpace Information (CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934

The United States National Distribution Centre does NOT hold stocks of AGARD publications.
Applications for copies should be made direct to the NASA Center for AeroSpace Information (CASI) at the address below.
Change of address requests should also go to CASI.

SALES AGENCIES

NASA Center for AeroSpace Information
(CASI)
800 Elkridge Landing Road
Linthicum Heights, MD 21090-2934
United States

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Requests for microfiches or photocopies of AGARD documents (including requests to CASI) should include the word 'AGARD' and the AGARD serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Note that AGARD Reports and Advisory Reports should be specified as AGARD-R-nnn and AGARD-AR-nnn, respectively. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Division
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements and Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic
Database or on CD-ROM)

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-836-0043-6